August 2020

Exploring Approaches to Effective Evaluation for Public Diplomacy Exchange Programs

Alison Opoku Donyina

Exploring Approaches to Effective Evaluation for Public Diplomacy Exchange Programs

Alison (Mason) Opoku Donyina

PIM 76

A capstone paper submitted in partial fulfillment of the requirements for a Master of Arts in Intercultural Service, Leadership and Management at SIT Graduate Institute in Brattleboro, Vermont, USA.

August 2020

Advisor: Dr. Bruce Dayton

CONSENT TO USE OF CAPSTONE

Abstract

This study explores what factors have limited strategic measurement of outcomes for '*mutual understanding*' across U.S. government-funded public diplomacy exchange programs by conducting a thorough desk review of available materials, completing a meta-review of twelve existing evaluation reports, and coupling this with interviews with eight evaluation practitioners. The sample of programs included were selected based on the criteria that *'mutual understanding'* was an explicit goal in the program design. Programs included: Fulbright Student Program, Fulbright Scholar Program, International Visitors Leadership Program (IVLP), Youth Exchange and Study Program (YES), Future Leaders Exchange (FLEX) Program, and the Peace Corps.

This paper synthesizes definitions of both *'public diplomacy'* and *'mutual understanding'* as used across the field of international relations and in government-sponsored people-to-people (P2P) programing. From a review of the literature, this study also captures posited frameworks, approaches and tools that can assist in measuring public diplomacy programming. It also reviews different approaches to evaluation and identifies strategies that have been used previously to measure mutual understanding and identifies pain-points for measurement of *'mutual understanding'* and provides some possible recommendations moving forward.

Key words: mutual understanding, public diplomacy, evaluation, measurement, soft power, Fulbright, IVLP, Peace Corps, Kirkpatrick Model, exchange program, mixed methods

Acknowledgements

This paper would not have been possible without the opportunity to work alongside the learning-oriented New Knowledge/Knology crew, the wise evaluators of the IIE Evaluation Unit, and the innovative Mandela Washington Fellowship team at IREX. Thank you for your collective wisdom and knowledge on both public diplomacy programming and evaluation. Thank you also to my advisor, Bruce Dayton, for his timely comments and helpful feedback.

Dedication

This paper is dedicated to my sweet Melody Aseda and midofo Kwaku. I’m so thankful for you both.

List of Abbreviations

ACPD	United States Advisory Commission on Public Diplomacy
AFS	AFS Intercultural Programs
America	United States of America
AMIDEAST	America-Mideast Educational and Training Services
APDI	Advancing Public Diplomacy Impact
BBG	Broadcasting Board of Governors
DOS	Department of State
ECA	Bureau of Educational and Cultural Affairs, Department of State
ED	Evaluation Division at the Bureau of Educational and Cultural Affairs
EMU	Department's Evaluation and Measurement Unit, Department of State
FLEX	Future Leaders Exchange Program
GAO	Government Accountability Office
G2G	Government to Government
G2P	Government to Publics
IIE	Institute of International Education
IPIC	Independent Practitioner Inquiry Capstone
IREX	International Research & Exchanges Board
IVLP	International Visitors Leadership Program
Meridian	Meridian International Center
NATO	The North Atlantic Treaty Organization
OMB	Office of Management and Budget
OSIRP	Office of Strategic Information, Research, and Planning, Peace Corps
P2P	People to People
PART	Program Assessment Rating Tool
PD	Public Diplomacy
PD-MAP	Public Diplomacy Model for the Assessment of Performance
PFP	Professional Fellows Program
PMI	Performance Measurement Initiative
RCT	Randomized-Control Trial
RPCV	Returned Peace Corps Volunteer
SNA	Social Network Analysis
U.S.	United States of America; United States
USAID	Agency for International Development
USIA	United States Information Agency
USG	United States Government
VRF	Volunteer Report Form for Peace Corps Volunteers
YES	Youth Exchange and Study Program
YFU	Youth for Understanding

Table of Contents

Introduction

This study examines approaches to evaluating U.S Government-funded 'people-to-people' exchange programs within the Peace Corps and the Department of State's Bureau of Educational and Cultural Affairs (ECA). It seeks to explore how components of programs designed to "increase mutual understanding" have been historically evaluated internally by ECA and Peace Corps staff and by external contracts. As public diplomacy is an entrenched foreign policy goal for the United States government, with significant funding[1] devoted towards such efforts (Epstein, Lawson, & Gill, 2018), it is important to both justify such funding and determine the impacts of these efforts. The framework of 'people-to-people' (P2P) is present within much of international exchange endeavors, including study abroad, but for the U.S. government there is an explicit goal of promoting "mutual understanding" between citizens of the United States and citizens of other nations.

People-to-people exchange, when sponsored by the public sector, is synonymous with the term citizen diplomacy and is based on the premise that individual interactions function as a means of establishing understanding between the individual and others. When these interactions are enabled through organized, state-sponsored programs, the individual is enlisted to play the role of an informal ambassador on behalf of their nation to the country that they visit. This act of facilitating understanding about their home country in the host country context is what is labeled people-to-people or citizen's diplomacy[2].

Programs within ECA[3], as well as the Peace Corps[4], target different populations and facilitate various activities, but there is one thing that all of these programs hold in common: In each, participants are considered ambassadors during their tenure abroad, each representing their country and fulfilling the hope

[1] $634.14 million was allotted for "Ed. & Cultural Exchanges" and $410.00 million for Peace Corps within the State Department, Foreign Operations, and Related Agencies Appropriations in the FY2017 Fiscal Budget (Epstein, Lawson, & Gill, 2018).
2 More information summarizing examples of citizen diplomacy can be found at the Center of Citizen Diplomacy website https://www.centerforcitizendiplomacy.org/about-us/understanding/
3 All of ECA's current programs can be found here: https://exchanges.state.gov/us/alphabetical-list-programs
4 Information about Peace Corps global presence and project sectors can be found here: https://www.peacecorps.gov/countries/

of their national government that their interactions can facilitate greater understanding about the countries that they call home. These programs have an implicit but still significant goal with outcomes that should be measured strategically in order to provide evidence that can inform contemporary foreign policy.

Exchange programs, which have become an entrenched ethos in U.S. foreign policy over the past 50 years, are given significant funding each year (Epstein, Lawson, & Gill, 2018). As these programs are ultimately paid for by American taxpayers, there is an expectation that the impacts stemming from participation in these programs will be measured through audits or evaluations to support justification for their funding. Historically, these programs have been evaluated on singular, irregular occasions. Additionally, the methodology used to determine impact for these programs has depended largely on who has undertaken the evaluation, typically either an internal team to the implementing organization, an evaluation unit of the sponsoring agency, or an external contractor, as evidenced by the various evaluation reports available for these programs. Across time and programs, evaluators have sought to measure the impact of these programs and ascertain impacts over time and across cohorts. Some evaluations have relied on monitoring data collected by program implementers and impacts have been reported largely based on participation, whereas others have been completed through external research with varying levels of rigor. As an illustration of on-going evaluation efforts for a single program, Fulbright Student has been evaluated by the Evaluation Division of ECA on several occasions, as well as by the implementing contractor, the Institute of International Education (IIE), and by external research teams on multiple occasions[5]. Each of these evaluative efforts have identified out various impacts that strengthen political resolve to continue the program's funding but are not easily combined into a portfolio of overall outcomes of the Fulbright Student program globally.

5 Evaluations commissioned by ECA for external research teams, including for Fulbright, can be found here: https://eca.state.gov/impact/evaluation-eca/evaluation-initiative/completed-evaluations. Evaluations conducted by the ECA Evaluation Division and IIE are not publicly available.

Across citizen diplomacy programming funded by the U.S. government, a review of previous evaluation efforts from sampled programs completed in this study demonstrates that over time there has been no specific approach or standardized method through which outcomes related to mutual understanding across these programs can be synthesized. Various attempts have been made through development of different frameworks or conducting different studies, but the tenure of such attempts is often influenced by ongoing challenges in sustaining evaluation consistent across the public sector. While the literature demonstrates that soft outcomes, such as attitudinal changes, are notably challenging to measure, considering the longstanding tenure of these programs, it appears there is a longstanding gap in measuring outcomes that support beliefs about the efficacy of public diplomacy. This study makes visible what has been measured and what is left unmeasured across a sample of evaluation approaches, in efforts to put considerations grounded in literature and lived experience that can inform future evaluations to better enable the measurement of "mutual understanding" for these programs.

Learning Objectives

As both a scholar and an evaluation practitioner, working to specifically to evaluate public diplomacy fellowship programming, this study holds both academic and professional weight. Through the process of conducting this Independent Practitioner Inquiry Capstone (IPIC), I sought to expand my knowledge and therefore improve my practice by accomplishing the following:

1) To understand the historical underpinnings of people-to-people diplomacy as part of the U.S. government's foreign policy and public diplomacy strategy;
2) To understand the umbrella goals and targets, as well as evaluation approaches and methodologies, used to evaluate the impacts of people-to-people exchange programs;
3) To familiarize myself with existing indicator frameworks previously used within evaluations across sectors that seek to measure various components relevant to people-to-people diplomacy;
4) To foster collaborative relationships with evaluation practitioners who are evaluating/have evaluated such programs; and
5) Reflect on my own practice as an evaluator.

Research Statement

This Independent Practitioner Inquiry Capstone (IPIC) explores what factors have limited strategic measurement of outcomes for '*mutual understanding*' across U.S. government-funded public diplomacy exchange programs. Strategic measurement is inclusive of both the strategy, the "how" of measuring outcomes for *mutual understanding*, but also the presence or lack thereof of an overarching strategy across a myriad of stakeholders to support the measurement of these outcomes. This research project seeks to: 1) explore and synthesize definitions of both *'public diplomacy'* across the field of international relations, and *'mutual understanding'* as used in government-sponsored people-to-people (P2P) programing; 2) review different approaches to evaluation used to determine impacts of such programming through conducting a meta-review of existing evaluation efforts; and 3) identify pain-points and opportunities for strategic measurement of *'mutual understanding'* moving forward.

The research focused specifically on these key programs: Fulbright Student Program, Fulbright Scholar Program, International Visitors Leadership Program (IVLP), Youth Exchange and Study Program (YES), Future Leaders Exchange (FLEX) Program, and the Peace Corps. This sample of programs was selected as "mutual understanding" is an inherent goal to the program design in response to government legislation.

To frame this inquiry, this study begins with a thorough review of the literature in defining public diplomacy, the historical context of exchanges, and existing theory and approaches to frame evaluation efforts to measure impacts of public diplomacy programming. Following is an explanation of the methodology, inclusive of the process of analyzing both collected data and evaluation reports, ethical considerations and limitations.

The main section of the paper comprises the research discoveries, which synthesizes information from a thorough desk review, meta-review of twelve evaluation reports, experiences of eight evaluators, and finishes with a conclusion and set of recommendations.

Literature Review

The span of literature on public diplomacy and international exchange is extensive; identifying gaps is only possible by completing a very thorough review from multiple angles. By first establishing how public diplomacy is commonly defined, especially in opposition to traditional diplomacy methods, then grounding the definition in dynamics of power and relevant international relations theory, the following literature review contextualizes international exchanges. By looking at the history, the context through which government-sponsored exchanges are both funded and implemented, and the evolving development of the case for evaluation for these programs, it concludes with an overview summarizing frameworks and approaches to evaluating of *'mutual understanding'*.

What is Public Diplomacy?

Coined by Edmund Gullion in 1965, "public diplomacy" was first codified as a field upon the establishment of the Edward R. Murrow Center of Public Diplomacy at Tufts University (Cull, 2009). However, the definition of the term has been somewhat murky. Scholars and practitioners have now concluded that public diplomacy entered the lexicon of 21st century diplomacy without any clear definition of what it is or how the tools it offers might best be used, as well as to whether public diplomacy's core a communications component takes place in a purely domestic or foreign sphere (Kelley, 2009). Public diplomacy is considered a dynamic, on-going process of cultivating connections than a product or message for distribution (Zaharna, 2009).

Common activities that fall under the umbrella of public diplomacy include those related to information, influence, and engagement. John Robert Kelley (2009) defines these as the following:

1. Information: information management and distribution with an emphasis on short-term events or crises;
2. Influence: longer-term persuasion campaigns aiming to effect attitudinal change amongst a target population (sometimes referred to as "moving the needle"); and

3. Engagement: building relationships, also over the long term, to cultivate trust and mutual understanding between peoples (be they groups, organizations, nations, etc.) (p. 73).

Though the targeted audience and content of public diplomacy activities is sometimes contested, public diplomacy, as defined by this paper, is understood as '*activities that influence public attitudes, both domestically and abroad, on the formation and execution of foreign policies'*. It encompasses dimensions of international relations seen as beyond traditional diplomacy, such as the cultivation of positive public opinion in other countries and facilitation of opportunities for interaction of private groups and interests from one country with another (Cull, 2009).

Traditional Diplomacy & the evolution of contemporary public diplomacy

Traditional diplomacy is typically thought of as government-to-government relations (G2G) (Snow, 2009), which leverages government officials such as the Secretary of State and/or Foreign Ministers in many countries to engage in interaction either via formalized meetings or highly publicized events. Traditional diplomacy is often bureaucratic and governed directly by the leader of a nation's foreign policy goals and enacted by their designated representative(s). Abroad, traditional diplomacy includes work by officially appointed Ambassadors and state-funded programming at Embassies that officially represent the interests of that nation on foreign soil. Public diplomacy was initially defined along these same bureaucratic lines as a form of influence through governments talking to global publics (G2P), which included efforts to inform, influence, and engage those publics in support of specific national objectives and foreign policies (Snow, 2009). Public diplomacy was descriptive of activities such as news broadcasting activities of the United States during the Cold War, which were ultimately one-directional processes through which one actor attempted to influence citizens who might in turn influence their governments (Cull, 2008).

Theories of public diplomacy have matured concurrently with evolving dynamics at play on the international stage. What was termed public diplomacy throughout the 20th century emerged as a result of two world wars and a competing balance of power between the Communist East and Capitalist West; 21st century public diplomacy now operates in a post 9/11 environment, which is dominated by a wholly different set of pressures: fractal globalization, preemptive military invasion, information and communication technologies that shrink time and distance like social media, and the rise of global non-state actors ranging from terror networks to bloggers (Snow & Taylor, 2009). Public diplomacy has now grown to include theory around how activities are seen to contribute to a nation's soft power by generating credibility, fostering values such as the belief in democracy, changing behavior, and increasing goodwill through activities including broadcasting and cultural diplomacy and exchanges (Pamment, 2014).

Contemporary public diplomacy must be comprehensive of the ways in which both government *and* private individuals and groups influence directly and indirectly foreign public attitudes and opinions that bear directly on another government's foreign policy decisions (Snow, 2009). *'People-to-people'* (P2P) diplomacy encompasses this version of public diplomacy. P2P diplomacy finds relationships as the core feature in such public diplomacy initiatives and focuses specifically on identifying and building relationships between individuals of different nationalities. These relationships are not intended as a means for enhancing individual national images or policies, but an end in itself. These relational initiatives seek to find commonalities or mutual interests between publics and then ways to link those publics via some form of direct interpersonal communication (Zaharna, 2009). As quoted by Zaharna (2009) Mark Leonard, London's Foreign Policy Centre, defines public diplomacy as "building relationships, starting from understanding other countries' needs, cultures and peoples and then looking for areas to make common cause" (p. 91). Building understanding can be done through specific actions or communicated messages but is also done effectively through symbols or artifacts (Zaharna, 2009). By understanding the extensive range of activities and relationships that fall under the P2P diplomacy umbrella, it illuminates the high level of complexity in trying to understand, or evaluate, overall progress towards public diplomacy goals.

The appeal of 'Soft Power' in response to anti-Americanism

In attempting to settle on a definition of public diplomacy, it begs the question: Why do nations need something to reach beyond methods of traditional diplomacy? Nancy Snow (2009), a leading scholar, answers this question succinctly: "Public diplomacy is inevitably linked to power" (p. 3). The type of power that Snow is referencing that is relevant to public diplomacy discourse is the idea of 'soft power'. Joseph Nye (2017) defined power as "the ability to affect others to get the outcomes one prefers, and that can be accomplished by coercion, payment, or attraction and persuasion. Soft power is the ability to obtain preferred outcomes by attraction rather than coercion or payment" (p. 1). Succinctly, enforcement of soft power, as opposed to hard power, results in getting others to appreciate you to such the extent that they change their behavior to your liking (Snow, 2009). A nation that wields soft power is able to convince other nations to support its objectives without having to employ their military, economic sanctions, or other coercive, intimidation-based methods of statecraft (Wyne, 2009). Hayden (2012) clarifies that the idea of leveraging soft power requires admit to some key theoretical assumptions about basic requirements for international persuasion, relationships between message and audience, and distinct implications of communication technology as a method of conducting of foreign policy.

Snow identifies some main tenants of a perceived soft power advantage:

> "What gives any country a soft power advantage is measured by several dimensions:
>
> 1. when culture and ideas match prevailing global norms;
>
> 2. when a nation has greater access to multiple communication channels that can influence, how issues are framed in global news media; and
>
> 3. when a country's credibility is enhanced by domestic and international behavior.
>
> The U.S. is at a comparative advantage with the first two and at a decisive disadvantage with the last dimension." (Snow, 2009, p. 4)

For Nye, soft power assumes that external, non-governmental actors are crucial, including individuals, NGOs, civil society, and the private sphere (Hayden, 2012). As international relations theory has continued to evolve, Hayden (2012) identifies that justifications for legislating soft power activities, including cultural exchange as a key method for public diplomacy, suggest that more traditional institutions of foreign policy are becoming less effective and further require instruments of soft power to maintain their ability to engage with foreign publics.

Thus, the philosophy of soft power is imbued directly into programmatic notions of public diplomacy. Hayden (2012) argues, “Public diplomacy is not the same thing as soft power, but the growth of public diplomacy programs justified in terms of soft power suggests a more significant shift… is taking place that merits further study” (p. 62). Contemporary public diplomacy focuses on approaches are saturated with the philosophy of soft power, expressed in the intention to attract multiple, often multi-national, stakeholders, with a promise of mutual gains (Kelley, 2009). Entrenched cultural narratives and practices in the U.S., holding tight to the importance of democratic values and the U.S. as a poster child for ideal values have deeply influenced how policymakers conceptualize enacting opportunities to exercise soft power. The choice to organize and fund programs like international exchanges derives directly from the value commitments of policymakers (Hayden, 2012). The notion of soft power is then ingrained and employed through resources like culture, values, and perceptions of policy legitimacy (Hayden, 2012).

As much of the United States’ interest has extended beyond the borders of countless sovereign nations over the years, a by-product has been a developing ‘anti-American’ sentiment. Though perhaps initially more a response to specific actions made by the U.S. in the foreign policy sphere, anti-Americanism has evolved into being understood as disdain for American culture and its system of values that combines respect for individual freedoms and pride in the capitalist ethic. As anti-Americanism is often understood simply as opposition to the enduring principles of American society, instead of America’s active conduct in the world (which does not always reflect those principles), exercising soft power to dismantle the sentiment at its heart is a fundamental shift (Wyne, 2009). With continually revolving sentiments of ‘anti-Americanism’ and compounded by a post-9/11 concern for national security, the global exercise soft power has become a mandate for multiple U.S. presidential administrations to date.

The method of choice to combat ‘anti-American’ sentiments abroad leveraging soft power is facilitating strategic exposure to Americans through the creation of formal, government-sponsored exchange programs, administered by third parties to been seen as more credible to other nations’ publics. Considering that many of these exchange programs are funded primarily by the U.S. government, citizen diplomacy has become a fundamental component of the United States’ approach to public diplomacy (Mueller, 2009). U.S. soft power is now synonymous with the term *‘mutual understanding’*, and by some in the public diplomacy sphere, it is believed that through the development of *‘mutual understanding’*, or familiarity with each other, citizens of the United States and other countries are more easily able to cultivate peaceful relations (Bean, 2017). By this definition, *‘mutual understanding’* refers to the idea that both a greater appreciation of other points of view on a mutual basis can contribute to a reduction in sources of conflict and a recognition of the universal benefits that stem from interdependence of global peoples and ongoing interchange of ideas, goods, and services (Scott-Smith, 2008). The above definition for *‘mutual understanding’* is adopted for the purpose of this study.

The version of soft power embedded into American exchange programs has also traditionally involved the weighting of one side of a supposed “mutual” understanding. While value is placed on Americans learning more about other countries, national security endeavors focus on leveraging soft power through emphasizing an increase in appreciation for the United States for visiting foreign nationals. By fostering exchanges, even in academia, the goal of *‘mutual understanding’* becomes synonymous with its hoped for results: a network of influencers abroad sympathetic to American interests due exposure that has resulted in familiarity and appreciation for the U.S. (Scott-Smith, 2008). By exporting Americanized knowledge through exchange, it also specifically ties these global influences into a network which positions institutions in the U.S. as the principal producer of knowledge as well as a gateway to credibility, professional growth and development (Scott-Smith, 2008). In its present iteration, U.S. public diplomacy operates on the assumption that the exposure of foreign audiences to U.S. values will lead to identification with American values, and that foreigners will perceive that U.S. values are universal and are in what they value too (Hayden, 2012). This assumption also holds that the most effective exposure is through personal interactions with Americans and America. U.S. public diplomacy is rooted in the idea of soft power "as a means by which a hegemon can retain its power in the international system" (Hayden, 2012, p. 64), whereby we will be able to "instill our values in other societies, remaking other cultures in our image" (Rider, 2015, p. 20).

Constructivist theory and influence on understanding outcomes of exchange

Another assumption crucial to understanding the use of exchange as a means to exercise soft power is that of constructivism. The theory of constructivism offers useful insights into and shapes strategies to capture when, why, and how ideas travel and change occurs (Scott-Smith, 2008, p. 186). Constructivism supports an iterative analysis of how interactions both produce and reproduce social structures while simultaneously shaping individuals' identities and interests. Scott-Smith frames the utility of this theory:

> What this theoretical field offers, therefore, is a means to situate the (potential) process of change that exchanges can initiate and, thereby, assess its political significance… The potential political reward from creating a situation whereby individuals may reconsider their identity, and so their interests, is obvious. And of all the options available in the public diplomacy toolbox that might achieve this, exchanges offer the best chance for success… (Scott-Smith, 2008, p. 185).

For constructivists, an individual's identity can transform when they experience a social situation that prompts them reconsider their self-conceptions, partnered with a perceived benefit to themselves or their communities. A constructivist model of public diplomacy is based on an assumption that norms, values, and identities in international relations are not defined by material power sources but are instead social constructs that can and will be influenced through exposure (Sevin, 2015). The theory is actualized for individuals who travel to the United States through an exchange program, are exposed to new ways of experiencing things through exposure to Americans and American institutions, and are received upon returning home with greater acknowledgement of their accomplishments. While this theory strongly informs political decision-makers' understandings of the value of public diplomacy, evidence (and the ongoing question of how to obtain it) as to impacts of public diplomacy is still under debate.

Cultural exchange's role in U.S. Public Diplomacy

With firm commitment to the efficacy of exchanges as a model for exercising soft power, former Under Secretary of State for Public Diplomacy and Public Affairs, Karen Hughes, was quoted: "Our education and exchange programs, I'm convinced, are the single most valuable public diplomacy tool." Quotes such as these by Department of State leadership and personnel continuously reiterate that international exchange is a fundamental building block of U.S. public diplomacy (Mueller, 2009, pp. 101-102). Even when considering the complex philosophies that encompass public diplomacy, international exchange remains appealing for predominant strains of international relations theory. Snow (2009) writes:

> In our rethinking public diplomacy, we will have to confront the two schools of thought that predominate, what have been characterized as the tender-minded versus tough-minded approaches…The tough-minded school is illustrated by the controversial firm Lincoln Group, whose website slogan, "Insight and Influence. Anywhere, Anytime," stands in sharp contrast to a mutual understanding approach. Signitzer and Coombs state that the tough-minded schools: hold that the purpose of public diplomacy is to exert an influence on attitudes of foreign audiences using persuasion and propaganda . . . The tender-minded school is illustrated by P2P and G2P strategies like the International Visitors Leadership Program and the Bureau of Educational and Cultural Affairs (ECA) of the U.S. Department of State, whose stated purpose is to foster mutual understanding between the people of the United States and the people of other countries around the world (Snow, 2009, p. 9).

The above quote illustrates both a key philosophical challenge of contextualizing educational and cultural exchanges as mechanisms of public diplomacy as well as an ongoing tension that has affected the administration and funding of these exchange programs within the government. For those of the "tough-minded" school or those that favor traditional diplomacy methods, over the years they have come to support exchanges as an extension and product of American propaganda efforts as elements in America's 'soft power.' Over time, even exchanges perceived as politically neutral have either political intent behind their creation or are promoted for the purpose of developing cross-border relations that should subsequently lead to political outcomes (Scott-Smith, 2009).

These exchange programs are often termed as opportunities for *'citizen diplomacy'*, which is the concept that the individual citizen possesses both the right and the responsibility to help shape U.S. foreign relations (Mueller, 2009). These citizen diplomats can be considered unofficial ambassadors for the U.S., where their role is to build individualized relationships that create trust, resulting in a collective web of connections that may later serve as the context for official dialogue and negotiations (Mueller, 2009). Considering the budget allocated annually, politicians perceive value in the opportunity to deputize unofficial ambassadors, who then may later serve in leadership roles in the U.S. and abroad. As an example, 97 current or former members of the U.S. Congress have participated in a Department of State sponsored exchange program (Evaluation Division of the Bureau of Educational and Cultural Affairs (ECA), n.d.) and as of 2018, more than 60 U.S. Ambassadors to countries around the world were returned Peace Corps Volunteers (RPCVs) (Coyne, 2018).

What exchanges then represent are a form of privatized international relations that plays different roles in U.S. foreign affairs, such as breaking down stereotypes and obstacles of cultural difference, transfer of knowledge and expertise, exposing those abroad to diversity in the U.S., as well as familiarizing Americans to the different views held beyond their borders (Scott-Smith, 2012). Comprised of an interchange of people, ideas, and opinions, the acquisition of global social contacts during exchanges holds much political power. Scott-Smith (2012,) states: “Exchanges contribute, in a loose, unguided way, towards an ever-denser global network society” (n.p.). The global, informal networks established hold major political importance (Scott-Smith, 2009), which is then clearly recognized by those who fulfill more traditional diplomatic roles. As an example, observed through responses to repeated surveys, U.S. Ambassadors have consistently ranked the International Visitor Leadership Program (IVLP) as one of the most valuable tools of U.S. public diplomacy at their disposal (Mueller, 2009). Snow (2009) summarizes the value of exchange succinctly:

> Educational and cultural affairs, where exchanges have their home, are the positive forces in foreign affairs. They work to unite the community of developed democracies; they assist the underdeveloped nations in becoming more viable and independent; they expand areas of mutual interest and mutual understanding; and they can help to build and strengthen global civic society (Snow, 2009, p. 235).

The history of sharing America

Framing the ongoing discourse around public diplomacy and international exchange requires grounding it in the overall history of international education in order to contextualize decisions that have been made by the U.S. government in earmarking funding for exchange programs.

William W. Hoffa (2007), a leading expert in the field of international education, identifies the origins of international exchange between Americans and other countries as something initially reserved for the elite upper class. The birth of international exchange in its more contemporary form took shape following the first World War in the first half of the 20th century, when ideas of building peace and understanding among nations by bringing together its possible future leaders initiated the founding of many international institutes and programs (Hoffa, 2007). Just a few short decades later, the devastation caused by World War II reignited many of these earlier beliefs with renewed zeal and exchange programming subsequently became a fixture in American international relations policy. Following World War II and throughout much of the Cold War period, between 1946 to 1991, is now considered the ‘Golden Age’ of international educational exchange (Snow, 2009). Hoffa and DePaul (2010) describe that the Cold War served as an important backdrop for all study abroad or educational and cultural exchanges as = geopolitical competition between the U.S. and the Soviet Union took place across many spheres: military, economic, political and cultural (p 17). A variety of global shifts, such as the Cold War, colonies declaring their independence from colonial powers and redrawing national boundaries, the civil rights movement, Cuban missile crisis, and President Kennedy’s assassination all mark the period of change occurring between the late 1940s and early 1960s (Hoffa & DePaul, 2010), which framed discussions and decisions for what would serve as strategic political investments for the future.

Much groundbreaking legislation was passed during these key decades, including a directive that from the 1948 U.S. Information and Educational Exchange Act (Public Law 402) to foster ‘mutual understanding.’ This law is commonly known as the Smith-Mundt Act for its two principal sponsors, Senator H. Alexander Smith (R-NJ) and Representative Karl E. Mundt (R-SD), and remains the legislative basis for America's foreign informational and cultural exchange programs to date (Snow, 2009). Another foundational piece of legislation, the Mutual Educational and Exchange Act or the Fulbright-Hays Act, was passed in 1961 (Hoffa & DePaul, 2010). The Fulbright-Hays Act updated and expanded provisions for the Fulbright exchange program initially made possible by the Smith-Mundt Act, through "financial and administrative upgrades, including earmarking new funding for grants, creating bi-national commissions for administration, increasing the number of participating countries, and seeking reciprocal economic support from partner governments" (Hoffa & DePaul, 2010, pp. 22-23). Hoffa (2007) argues that the "main thrust of the 1961 act was to tie together and simplify a disparate array of existing governmentally-supported international cultural programs" (p. 23).

But the methods of leveraging soft power were honed during the Kennedy Administration. While President Kennedy served in the early 1960s, a variety of international shifts occurred, and the purpose of public diplomacy consolidated under a banner of 'winning hearts and minds' (Armstrong, 2009). Around this time, the American public grew more aware of anti-American sentiment. An example is the influential book The Ugly American, published in 1958, which encouraged Americans to interact meaningfully with citizens of other countries by learning about and listening to them (Hoffa & DePaul, 2010) to combat negative perceptions abroad. The philosophy of soft power and the potential of exchanges as a meaningful two-directional approach to public diplomacy, opening up spaces for dialogue and exchange of alternative viewpoints, became a national priority under the catchphrase *'mutual understanding'* (Scott-Smith, 2009).

The oversight and administration of international exchanges has also changed hands over-time within the U.S. government's infrastructure. Public diplomacy initiatives were housed for many years in an agency called the U.S. Information Agency (USIA), which was formed in 1953, with its role to explain American values and further U.S. national interests through overseas information programs, as well as to promote mutual understanding enabling educational and cultural activities around the world (Duffey, 2009). ECA, the designated department for administering and facilitating contracts with partners to implement many of these exchange programs was originally housed in USIA but was later transferred under the purview of the DOS's public diplomacy agenda in 1999. ECA's mission has remained the same: to foster "mutual understanding between the people of the United States and the people of other countries" (Bureau of Educational and Cultural Affairs (ECA), n.d.), and this mission of strengthening mutual understanding is a legislatively mandated objective of all U.S. exchange programs (Banks, 2011).

Now, for over 60 years, U.S. public diplomacy has placed a high priority on cultural and educational exchange activities across a continuum that includes both hands-off and hands-on approaches. Mueller summarizes this value:

> Youth exchanges, such as those sponsored by the Experiment in International Living of World Learning, AFS Intercultural Programs, Youth for Understanding, and an array of other organizations (both nonprofit and for profit) constitute a critically important category of exchange programs. Although not a part of official U.S. public diplomacy activities, these programs certainly supplement public diplomacy. They enable participants to engage in vitally important efforts to build constructive relationships on behalf of Americans with people around the globe (Mueller, 2009, p. 106).

In the past two decades, the events of September 11, 2001 led to new, targeted interest in international exchanges and their importance to national security and foreign policy objectives, especially in the Arab world (Snow, 2009). As a result, all education abroad, but specifically educational or cultural exchanges, began to merge with broader priorities like enhancing national security in the latter part of the twentieth century (Hoffa & DePaul, 2010). One professor from University of Albany summarizes some of these shifts:

> The aftermath of 9/11 created a hyper-interest in the area of Middle East studies and the Arabic language, including one of the more significant booms in the hiring of faculty in those areas… However, the subsequent "War on Terror" provided new funding streams for academic programs the focused on national security interests and a handful of new academic fields were launched as a result. For example, we have seen the rise of programs dealing with cyber-terrorism, information security, and disaster response (Media Relations Office - University of Albany, 2011).

The new funding streams saw the launch of targeted public diplomacy programs such as the YES program for students in countries with significant Muslim populations to study in the U.S. for a year (Kennedy-Lugar Youth Exchange Program, n.d.).

Yet despite the targeted spending and high value placed on the value of exchanges in furthering America's soft power, there remains a large gap in knowledge regarding the long-term impacts of U.S. government financial investment in this method of public diplomacy over the past 50 years. Carissa Gonzales (2015), after conducting dozens of personal interviews with public diplomacy leaders, concluded "Rigorous evaluation of our programs is crucial in order for interagency policymakers and Department leadership to see Public Diplomacy for what it actually is: an instrument of national power" (p. 37).

Complexity, context, and the case for evaluation

Exchanges are traditionally contracted out to private sector operators to demonstrate the "apparent political neutrality" of these exchange programs (Scott-Smith, 2008, p. 182). An example for IVLP is illustrated in the quote below:

> While the State Department remains responsible for the overall functioning of the IVLP [and other similar exchange programs], it contracts out the organization of participants' itineraries to private sector operators such as the Institute for International Education and Meridian. This hands-off approach by the U.S. government raised the credibility of the program and enabled grantees to testify to the lack of political interference in their experience when they returned home (Scott-Smith, 2008, pp. 182-183).

Two contextual factors that further compound challenges in identifying impacts of exchange are the decentralization of the public sector and competitiveness across the social sector for funding. In today's public funding environment, with increased pressure to spend tax-payer money appropriately, government agencies no longer award contracts to providers because of what they are as non-profits, but instead what they can do and how efficiently and affordably they can do it (Eikenberry & Kluver, 2004).

Separately, with the emergence of private sector philanthropists as an alternative source of funding for non-profit partners, requirements for such funding are then linked to a market-based approach, where requirements for grants incorporate application of venture capital principles and practices to enable social change (Eikenberry & Kluver, 2004). Private funding sources expect a ROI (return on investment), a SROI (social return on investment), FROI (financial return on investment), or EROI (emotional return on investment) as a result of their funding (Eikenberry & Kluver, 2004). For most social sector organizations dependent on a revenue stream to ensure its function, donors (private or public) assume a central focus and donor priorities are quickly adopted as the organization's own (Weerawardena, McDonald, & Mort, 2010). The market-based paradigm with donors' needs at the center has seamlessly been adopted by the public sector and forms the basis for which contracts to implement international exchanges on behalf of the State Department are designed.

The environment fostered by contracting out the implementation of exchange programs has become increasingly crowded in the private and non-profit sectors. Competition for funding, in addition to the pressure to constantly improve delivery and prove value of services provided, has resulted in the reality that across sectors, organizations, and regions, similar programming is being delivered by competing contractors who seek to maintain their funding rather than sharing best practices externally. Cultural and educational exchange programs, such as those funded by ECA, have been distributed among numerous contractors who have won a bidding war for these funding contracts. These independent contractors, for-profit or non-profit, hold responsibility for delivery of these programs, and there is little opportunity for collaboration across due to the continual competition for contracts and funding.

The lack of collaboration described among contractors extends to evaluation practice as well. This competition and demand for quick, measurable results has put significant pressure on implementing organizations and has limited efforts to capture longitudinal impacts: Banks (2012) succinctly describes the challenge:

> There is pressure from funders for quick results. Telling last year's million-dollar donor to wait ten years to see results will likely ensure that additional funds will not be forthcoming. This pressure to see a quick return on investment can compel agencies to put more resources toward measuring short rather than long-term impact (pp. 31-32).

However, Hayden (2012) reminds that the influence of soft power can be identified through careful process tracing and outcome mapping to identify how an international actor translates resources into effective outcomes. Thus, the case for evaluation – by capturing outcomes for exchange programs, it is possible to illustrate the extent to which instruments of soft power, namely exchanges, are influencing global attitudes, perceptions and relationships with the U.S. A requirement of ECA's exchange program solicitations contracts has sometimes included references to establishing monitoring and evaluation for the program by the implementing partner, but until recently, ECA had not required a standard approach as to how evaluations were to be done or even what should be measured for these types of programs. Even when evaluation has been done as a condition of funding, they are often limited to measuring outputs of programs, rather than outcomes or impacts. Peace Corps[6] also only recently adopted a more standardized approach to evaluation across its many posts, yet still distinctly lacks ways to measure public diplomacy goals. In both agencies, there are specific departments intended to oversee evaluation[7], but often they function as third-party evaluators instead of working to synthesize findings or manage knowledge across programs. Consequently, there is little consensus in the approach to measure impact across the larger body of these government-sponsored citizen diplomacy exchange programs, though the overall program goals all align.

What the literature does demonstrate is that there have been some efforts to gather evidence to determine the value of these exchanges in breaking down stereotypes, but there exists a distinct critique that exchanges should be structured to achieve maximum impact while being more diligent about measurement (Bellamy & Weinberg, 2008). Barriers to measuring impact are concluded to be as much organizational and operational as financial and many stem from the challenges working across a dispersed community of federal agencies, NGOs, and academic institutions (Bellamy & Weinberg, 2008). This context limits any attempt to align efforts around a common set of objectives and standards for the evaluation of the portfolio of exchange programs (Bellamy & Weinberg, 2008). Former Ambassador

6 In the Peace Corps 2018-2022 Strategic Plan, they initiated the development of logic models for all different sector projects across countries for the first time (https://files.peacecorps.gov/documents/open-government/pc_strategic_plan_2018-2022-annual_plan_2019.pdf)

[7] ECA has a specific Evaluation Division (https://eca.state.gov/impact/evaluation-eca) and Peace Corps utilizes a decentralized strategy that includes embedded Evaluation staff for each region and at each post, the Office of Inspector General and the Office of Strategic Information, Research, and Planning within the agency (https://www.peacecorps.gov/about/inspector-general/reports/).

Christopher Ross described exchanges as “a longer-term effort to develop an overseas understanding and appreciation of U.S. society—the people and values of the United States…Success on the information front can be measured. In contrast, gauging the success of exchange programs is more intangible and requires time and patience” (Kelley, 2009, p. 74).

Considering the long history, the scale of exchange alumni populations, breadth of global stakeholders and sheer diversity of activities undertaken under both the banner of ECA and Peace Corps-sponsored programming in over 160 countries, codifying the overall increases in “mutual understanding” with any rigor proves near impossible. Bettie sums this up in saying:

> The value of exchange of persons programmes is not always obvious. It is difficult to measure intangible results like an increase in mutual understanding, or to assign credit for an individual’s change in attitude. There are many factors that can influence our views of foreign nations and cultures. While returned grantees express positive attitudes of their hosts, quantifying such outcomes and linking causality to the exchange programme is fraught with complications (Bettie, 2015, 359).

After over 80 years of State Department funded exchanges and nearly 60 years of Peace Corps, there are enough surveys and anecdotes to support the view that those who participate in the various exchange programs regard it as a positive experience (Scott-Smith, 2012). However, entrusting America’s image to individuals can have negative repercussions. As decisions about proving the value of exchanges intentionally highlight the positives on exchange, very few sources acknowledge any negatives, such as experiences of racism or discrimination for international visitors to the U.S. as well as discouraging interactions with Americans abroad. There is very little publicly available across the literature about unintended consequences of exchanges and challenges that individuals’ experience upon their reintegration home, be it loss of employment or challenges with their social support systems.

Despite historically challenging measurement, Giles Scott-Smith (2008) clearly outlines the case to push forward and implement stronger evaluative methods moving forward.

> Evaluation studies conducted intermittently during the past fifty years by the State Department and contract agencies have provided considerable anecdotal evidence indicating the favorable outcomes that these activities have generated. Yet a gap remains between the usefulness of such qualitative research, with its vagaries of human judgment, self-perception, and memory,

> and the demands of social science for empirically verifiable research, fact, and proof… there are political effects from exchanges, which, although often fragmentary and inconsistent, can be identified and understood in a coherent framework (p. 174).

As much of public funding and policy has become driven by evaluative findings, there appears to be a cause, supported by the literature, to more strategically measure and evaluate public diplomacy-specific outcomes of the Peace Corps and cultural and educational exchange programs in order to determine global impacts that can better shape foreign policy moving forward. Categories that have emerged as being worthy of study include personal and professional networks, attitudinal shifts towards the U.S. by foreigners and towards other countries by Americans, as well as both positive and negative outcomes stemming from participating in an international exchange.

Pahlavi (2007) summarizes a cornerstone of the argument to improve and align evaluation practice:

> Successful demonstration of the efficacy of public diplomacy would be a powerful argument… that the pursuit of interests may be undertaken with means other than brute force. The theorists of soft power have until now held to vague propositions that, although seductive, have never really been ascertained. Progresses in this domain should allow us to make good this void and push the subject further by demonstrating concrete uses of soft power. A performing evaluatory system would reveal the increased complexity of the nature of state power within the international system. We would be able, in other words, to confirm the seductive hypothesis, still unverified, that soft power 'matters' and 'can have hard results' (pp. 279-280).

Measurement approaches to evaluating impacts of public diplomacy

Across the field of public diplomacy scholarship, there is some consensus about ways to measure soft power. Joseph Nye, Jr. (2006) suggests measuring and comparing the cultural, communications, and diplomatic resources that can produce soft power for a country. A strategy that is often undertaken to measure and draw conclusions about the activities that produce soft power for a country is evaluation.

Evaluation has been defined by public relations scholars as a “form of research that seeks to determine the relative effectiveness of a program by measuring program outcomes such as changes in the level of awareness, understanding, attitudes, opinions and/or behaviors of a certain group against a predetermined set of objectives” (Pahlavi, 2007, p. 22). A father in the evaluation field, Michael Quinn Patton, defines it as “the systematic collection of information about the activities, characteristics, and outcomes of programs, for use by people to reduce uncertainties, improve effectiveness, and make decisions" (Patton, 2008, p 39). Watson and Noble (2007) summarize Noble's seven principles of evaluation for public relations campaigns, which often mirror public diplomacy initiatives:

1. Evaluation is a research-based discipline: Its purpose is to inform and clarify and it operates to high standards of rigour and logic…
2. Evaluation looks both ways. Evaluation is a proactive, forward looking and formative activity that provides feedback to enhance programme management. It is also a reviewing, backward looking summative activity that assesses the final outcome of the campaign/programme…
3. Evaluation is user and situation dependent. Evaluation should be undertaken according to the objectives and criteria that are relevant to the organization and campaign concerned…
4. Evaluation is short term. Short-term evaluation is usually campaign or project based... Short term in this context definitely means less than 12 months.
5. Evaluation is long term. Long-term evaluation operates at a broader, strategic level and usually concerns issues management, corporate reputation, and/or brand positioning… long term, regular feedback from evaluation research can help fine-tune planning and implementation as well as measuring results.
6. Evaluation is comparative. Evaluation frequently makes no absolute judgements but instead draws comparative conclusions.
7. Evaluation is multifaceted. Public relations has been established as a multi-step process, if only because of the additional stepping stone represented by the media. A range of different evaluation methodologies are required at each step (or level), with process evaluation, for example, being used to enhance the effectiveness of impact effects (Pahlavi, 2007, pp. 23-24).

There are two types of evaluations that are frequently employed for public diplomacy program evaluations: Process evaluations and impact evaluations. Process evaluations tend to look forward and provide feedback and are frequently short-term efforts. Impact evaluations, on the other hand, look at outcomes at different levels and over a longer period of time. Banks (2011) summarizes, in his Resource Guide to Public Diplomacy Evaluation, that a process evaluation is intended to identify how well a program is working while an impact evaluation:

> An impact evaluation… tries to get at the "why" of the program. Why did we conduct it in the first place? What were we trying to accomplish, and did we succeed? It asks questions designed to determine whether the PD program had produced the desired impact, usually defined as a positive change in awareness, attitude, or behavior. Impact evaluations are more difficult and time-consuming to design and conduct, may require pre-during-and post-program assessments, and demand a significantly higher level of analysis (Banks, 2011, p. 29).

Yet outcomes as nuanced as attitudinal changes towards the United States and increased awareness of other countries are notoriously hard to measure. Evaluators have proposed a variety of frameworks and approaches to support trying to determine if public diplomacy programs are succeeding.

Frameworks, models, and approaches

For evaluators, there are numerous the different models, approaches, and frameworks through which to attempt to identify impacts. For outcomes stemming from programs like cultural exchanges and trying to understand their impacts in relation to public diplomacy, the literature demonstrates there are some tailored frameworks and approaches that can support an evaluation study. The variety of frameworks to illustrate different approaches to determining impact for public diplomacy activities, such as those exemplified in ECA's cultural exchanges and Peace Corps volunteering efforts. They are summarized throughout this section.

Soft Power 30 Index

One specific tool that has emerged in the past few years is the Soft Power 30 Index, which was developed through a partnership with the University of Southern California (USC) Center on Public Diplomacy (CPD) and Portland, a strategic communications consultancy firm. The Soft Power 30 Index combines objective data and international polling to illustrates the potential for influence among a set of 30 countries (McClory, 2017). The Matrix focuses on several subthemes, including categories such as Government, Culture, Education, Global Engagement, Enterprise, and Digital and measures across more than 75 metrics, gathered through data that is observed and sourced through international polling and allows for an overall ranking of global soft power resources (McClory, 2017). One missing element from the Index is that of cultural exchanges.

NATO's Framework for the Strategic Planning and Evaluation of Public Diplomacy

The North Atlantic Treaty Organization (NATO) has made efforts to framing guidance around evaluating public diplomacy initiatives. In their Framework for the Strategic Planning and Evaluation of

Public Diplomacy, they recommend the following methods and tools to evaluate the effects for P2P diplomacy efforts:

Table 1. Methods and tools for Evaluation Public Diplomacy Activities in Bliss, et. al. (2013).

	METHODS & TOOLS		
	Formative Evaluation	**Output Evaluation**	**Impact/Outcome Evaluation**
Data Collection Methods and Tools	• Desk research • Face-to-face or telephone interviews with opinion formers • Online interviews • Traditional focus groups • Online focus groups • Online panels of opinion formers • Narrative inquiry • Social media monitoring and analysis	• Observation • Live feedback during a conference (App or web based) linked to an online survey platform • Conference exit polls • Follow-up online surveys • Narrative inquiry • Social media monitoring	• Semi-structured interviews • Online interviews • Online panels of opinion formers • Online surveys • Omnibus polls • Narrative inquiry • Social media monitoring • Media content analysis

Note: Reprinted from Table 9 in Bliss,K, & et. al. (2013). A Framework for the Strategic Planning & Evaluation of Public Diplomacy, NATO, Joint Analysis and Lessons Learned Centre: Portugal, p.41.

For NATO, people-to-people engagement is a key aspect of outreach which enables the Alliance to build relationships with individuals, influencers, and target audiences, so their proposed measurement practice is designed to look at the number of people reached and of individuals who keep in contact with NATO after a public diplomacy project (Sevin, 2015). Sevin (2015) observes that this relationship-based approach leads many to utilize social network analysis as a method to measure the impacts of public diplomacy interactions.

Pathways of Connection framework

Another proposed approach to measure the impacts of diplomacy is the Pathways of Connection framework (Sevin, 2015). This framework is intended to inform analysis at the level of six pathways and proposes the likelihood to observe impact in the realms of public opinion, relationship dynamics, or public debates (Sevin, 2015), and to therefore focus efforts of measurement in identifying shifts in those categories. The below image illustrates the model in Figure 1:

Figure 1. Six pathways of connection (Sevin 2015).

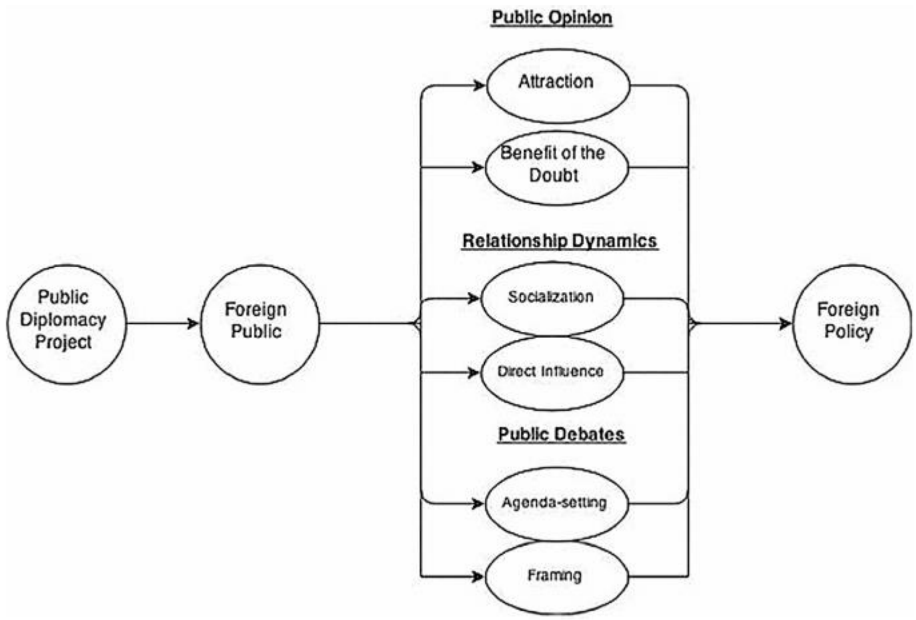


Figure 1. Image of the Six pathways of connection. Reprinted from Sevin, E. (2015). Pathways of connection: An analytical approach to the impacts of public diplomacy. Public Relations Review, p. 567.

Sevin (2017) further expounds that by tracking outcomes through these three layers, public opinion, relationship dynamics, and public debate, practitioners and scholars are most likely to observe the impact of any public diplomacy projects. These layers can also be viewed as "areas of impact" (Sevin 2017). The six-pathways of connection framework provides a way to evaluate outcomes stemming from public diplomacy programming through ensuring criterion for success and evaluation are set, a multilayered approach is able to summarize the outcomes of projects, and the process-based application of this framework makes it possible to argue for the causality between the projects and the outcomes (Sevin 2017).

The British Council's 3 Metric approach

The British Council and its Head of Evaluation, Ian Thomas, has also put forth some approaches towards how to measure soft power. Thomas (2018) writes:

> The British Council approach is based on mutuality and co-creation across its cultural relations programs operating in 110 countries… looking at how the British Council might approach evaluating the arts and soft power, considers soft power a dynamic process. However, the vagueness of the concept has limited its effective deployment. Evaluation of soft power suffers

> from the absence of clear objectives or overly generalized objectives. There is difficulty in identifying and isolating the "object of study," developing meaningful indicators of success and agreeing on routes to evaluation. The evaluation of soft power effectiveness should be two directional: looking at both the delivery agents' resources, capabilities and behaviors along with the receiving audience's perceptions, and behaviors toward the soft power delivery agent. Soft power can too easily become an empty phrase. In reality what it means is the cumulative, long-term bank of national assets—moral and cultural, which predispose people to listen to a country (n.p.).

In the British Council's approach to defining soft power as a goal, they use a framework in which they utilize a 360-degree integrated approach based on three metrics: 1) Return on influence; 2) Return on relationship; and 3) Return on investment (Thomas, 2018). Thomas (2018) describe leveraging a combination of approaches, including digital sentiment analysis tied to hashtags, stakeholder interviews, network and media analysis, and assessment of business or economic activities tied to their programs. An additional model they use is called the Cultural Value Model, developed by the Open University, which supports looking at cultural value from multi-stakeholder perspectives. The Cultural Value Model intends to shift the frame of analysis from impact to value, leveraging a more engaged, participatory approach to performance evaluation that is still viewed as robust and rigorous (Bell, Gillespie, & Wilding, 2016).

The Kirkpatrick model

An additional framework for determining how to evaluate public diplomacy programs is a revised model of Donald Kirkpatrick's Four Levels of Evaluation. Martel (2018) describes

> Kirkpatrick's model describes the levels of impact that measure change resulting from an academic experience, ranging from a short-term training to a full-degree program. The model outlines levels of change starting from the individual and proceeds to measure change at the institutional level… Kirkpatrick's methodology is useful in that it de-emphasizes the individual as the only possible change outcome. Many studies end at Kirkpatrick's level one or two, assessing the impact solely at the individual level. Kirkpatrick, rather, focuses on the application and behavioral transfer of knowledge to one's environment or secondary beneficiaries. Further, Kirkpatrick's model is goal based, meaning that the model identifies the intended goals of the intervention, but does not necessarily evaluate the processes to achieve those goals" (Martel, 2018, pp. 289-291).

Martel's adapted illustration of the Kirkpatrick model is depicted below:

Figure 2. Adapted Kirkpatrick Model (Martel, 2018)

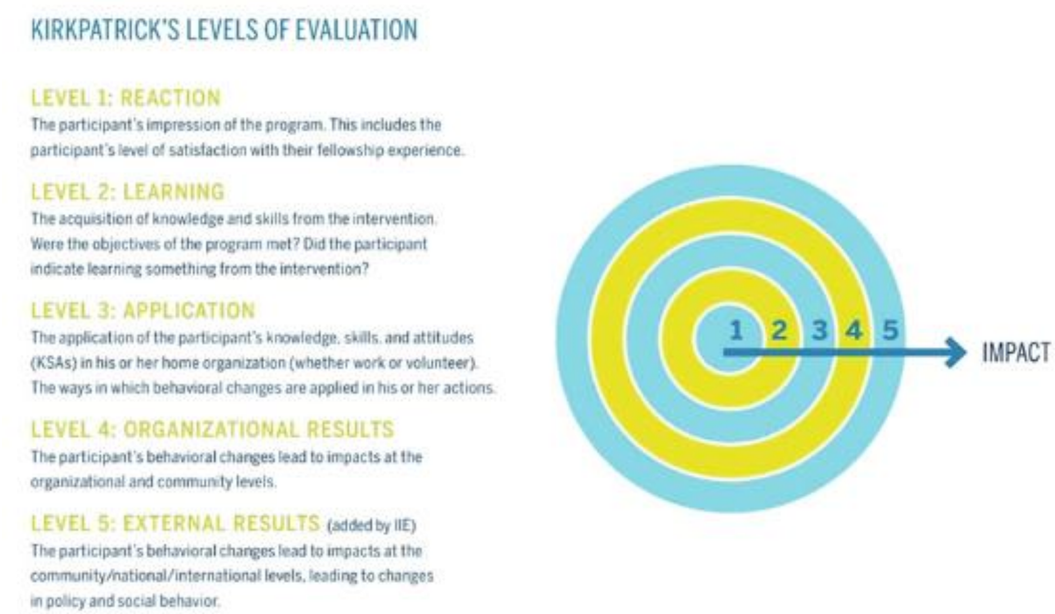


Figure 14.2. Image of A revised version of the Kirkpatrick model focuses on two levels of potential impact beyond the individual: organizational and external (societal). Reprinted from Martel, M. (2018). Tracing the spark that lights a flame: A review of methodologies to measure the outcomes of international scholarships. In J. e. Dassin, International Scholarships in Higher Education (pp. 281-304). Chan: Palmgrave Millan, p. 290.

The dialogic model of public diplomacy

Another proposed model is the dialogic model of public diplomacy by Kathy Fitzgerald (2011). Based on the theory of dialogue as more than a communication process or strategy but as a method to identify attitudes with which individuals approach each other (Fitzpatrick, 2011). This theory frames a dialogic orientation as one that is towards mutual understanding rather than self-interest (Fitzpatrick, 2011). Her model includes eight criteria, described in the table below, and serves as a framework that can support evaluating public diplomacy policies and practices globally:

Table 2. Criteria for a Dialogic Model of Public Diplomacy by Fitzgerald (2011).

Criteria	Requirements
1. Mutuality	Mutuality requires reciprocity of parties and interests, as well as the opportunity for expression.
2. Presence	Presence requires that parties be available and open to each other and involved in matters that affect them "in the present."

3. Commitment	Commitment requires that parties be proactive in engaging with others and be willing to participate in efforts to reach mutually satisfying outcomes.
4. Authenticity	Authenticity requires a presumption of honesty, transparency and genuineness by each party.
5. Trust	Trust requires that authority and power be set aside and that each party be empathetic to the other.
6. Respect	Respect requires parties to recognize and accept “strange otherness,” or the unfamiliar views and unique traits of others.
7. Collaboration	Collaboration requires sincere engagement between parties in which the relationship is not viewed in terms of winning or losing or as an attempt to defeat the other’s ideas.
8. Risk	Risk requires that parties accept the uncertainty of dialogic outcomes.

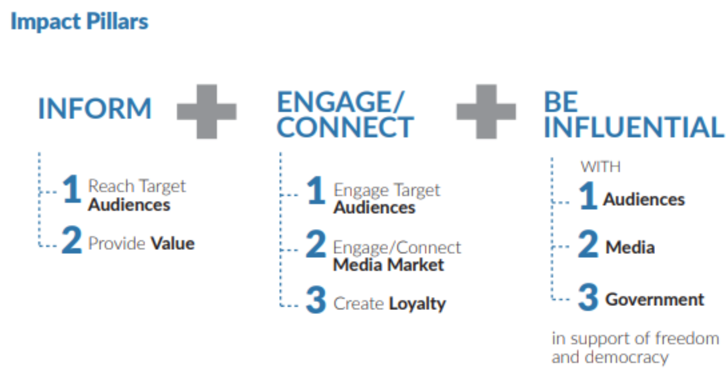
Note: Reprinted from Table 2: Criteria for a Dialogic Model of Public Diplomacy in Fitzpatrick, K.R. (2011). U.S. Public Diplomacy in a Post-9/11 World: From Messaging to Mutuality. Figueroa Press, Los Angeles, p. 21-22.

BBG’s Impact Model

The Broadcasting Board of Governors (BBG) is America’s civilian international media agency has developed a customized Impact Model (Broadcasting Board of Governors (BBG), 2018). This a model leveraged to measuring impact in the varied and complicated media environments in which BBG networks operate and looks beyond sheer audience size to assess any concrete change that the news and information provided by their programming has made in the lives of audience members, in the local media sector, and among governments (Broadcasting Board of Governors (BBG), 2018).

The model frames strategic questions to support identifying reach, or number of individuals accessing their materials, as well as the perceived value. The next aspect of impact is the multiplier effect – how individuals who are consuming media produced by these platforms are then being broadcast to other audiences and trusted by foreign publics. The final aspect of impact is the idea of influence, from either the reach of their message being received at key levels or in shifting opinions of individuals or flows of information for consumers.

Figure 3. BBG's Impact Model



Note: Reprinted from Appendix: Impact Model in Broadcasting Board of Governors (BBG). (2018). BBG Strategic Plan 2018-2022: Information Matters: Impact and Agility in U.S. International Media. Washington D.C.: Broadcasting Board of Governors (BBG), p 31.

Public Diplomacy Model for the Assessment of Performance (PD-MAP)

Authored by a research team with The University of Texas at Austin, the PD-MAP was developed to serve as a comprehensive and flexible measurement tool that allows an evaluator to quantify the results of public diplomacy programs and evaluate success in meeting three strategic defined outcomes of U.S. public diplomacy programming (Matwiczak, 2010). The three goals include increasing understanding of US policy and culture, increasing favorable opinion towards the U.S. and increasing the U.S.'s influence in the world at large (Matwiczak, 2010). The PD-MAP not only provides a key framework in identifying the target audiences of public diplomacy programs as government stakeholders, elites and general publics, but also defines outcomes and suggested measurements. Main outcomes framed in the PD-MAP are categorized under three buckets: understanding, favorability, and influence and outlines a comprehensive list of corresponding indicators to support measuring these outcomes across the different audiences (Matwiczak, 2010)

Four public diplomacy evaluation models

Pamment (2014) also explores ways to approach public diplomacy (PD) evaluation. His view is that there should be interpretation of evaluation practice based on the model of evaluation (Pamment, 2014).

He identifies four models that encompass approaches to evaluating public diplomacy programs, each with different methods, theories that inform design, and the anticipated results from the evaluation efforts in the table below.

Table 3. Overview of PD Evaluation Models by Pamment (2014).

Articulation	Methods	Theory of influence	Anticipated results
Output models	Ad hoc, Press clippings, AVE, OTS	PD as outputs	Proof of labor/reach/volume
Output models	Logic models, Impact measurements	Soft power = hard effects	Proof organization is effective/efficient
Output models	Surveys, Attitudes, Favorability	Reputation management	Proof of influence over ideas & values
Output models	Hubs & Multipliers, Forming alliances	Relationship management	Proof of attention to relationships & other perspectives

Note: Reprinted from Table 1: Overview of PD evaluation models in Pamment, J. (2014). Articulating influence: Toward a research agenda for interpreting the evaluation of soft power, public diplomacy and nation brands. Public Relations Review, p. 58.

For public diplomacy activities, it is possible to look at four different levels of outputs by codifying an appropriate theory of influence. For Pamment (2014), the theory of influence illustrates the core intent of the program design. For a program that intends to foster deep relationships, one particular output model would be more appropriate to use as a framework for evaluating that program; for one that intends to bolster the credibility of an individual or foster alignment to democratic values, a different model, informed by its theory of influence, would be more apt.

Elite comparison approach

Another approach to public diplomacy evaluation that sought to measure the aggregate impact of public diplomacy on foreign publics in a look across different programs was the Advancing Public Diplomacy Impact (APDI) initiative, which was launched in 2007 by the State Department's Evaluation and Measurement Unit (EMU) (Banks, 2011, p. 28). This initiative focused on surveying the differences in attitudes between foreigners who had participated in USG-funded public diplomacy program within the a five year time range, and also surveyed a control group of elites from the same demographic categories who did not participate (Banks, 2011). In 2007, this initiative successfully surveyed 1,800 elites in eight locations from multiple sectors including journalism, the arts, NGOs, academia, and politics and two years later in 2009, they conducted a similar study but increased the sample size to 6,500 (Banks, 2011). The survey focused on determining increased understanding of U.S. society, values and policies, greater favorability toward the U.S., changes in attitude toward U.S. policies and influence globally, as well as program satisfaction and receptivity to future engagement (Banks, 2011). Banks (2011) summarizes the study and relates it to the greater challenge in saying:

> APDI findings suggested that repeated PD engagement over time can have a positive impact on participants' overall views of the U.S. The areas where PD engagement seemed to produce the weakest results were in support for U.S. foreign policy and views toward U.S. global influence. These results mirror the findings from numerous past evaluations of exchange programs, where positive changes in attitude toward U.S. foreign policy stood out as the hardest objective to achieve. That said, the results in these two areas among PD program participants still surpassed those for non-participants. One of the challenging issues with APDI, as with exchange evaluations, is trying to determine the degree to which participants come to the PD program pre-disposed to sympathize with its objectives and/or its sponsors (Banks, 2011, p. 28).

In leveraging an approach to collected data in a comparison study with a large sample size, this study attempted to capture effects specific to participating in a public diplomacy program. However, there were still noted limitations.

Existing meta-reviews on evaluation related to public diplomacy

While the APDI's study was done within government, two other studies attempt to look at the overall outcomes, done by the Institute of International Education (IIE), long-time third-party implementer of many ECA-funded exchange programs (most notably the suite of flagship Fulbright programming). One of these reports was completed in 2011 and another in 2019. The study in 2011 conducted a meta-review by identifying twelve different evaluations and framing them as case studies, which then provided some overarching recommendations for improving measurement of public diplomacy programming (Bhandari & Belyavina, 2011). The study identified a sample that included high school exchange programs, international volunteer programs, and professional and leadership exchange programs (Bhandari & Belyavina, 2011). The study framed the context of citizen diplomacy and evaluation work done to date, identifying that often evaluations have focused on measuring short-term results on individuals and on communities, and commonly use participant and beneficiary constituent surveys, interviews, and focus groups (Bhandari & Belyavina, 2011). In their meta-review, they identified some key outcomes related to exchange programming as revealed through their sample of twelve evaluations, depicted below:

Figure 4. Key Outcomes- Comparison of U.S. and International Participant Outcomes (Bhandari & Belyavina, 2011)

Image of Figure 1: Key Outcomes- Comparison of U.S. and International Participant Outcomes. Reprinted from Bhandari, R., & Belyavina, R. (2011). Evaluating and Measuring the Impact of Citizen Diplomacy: Current Status and Future Directions. New York: Institute of International Education, p. 13.

In the 2019 study, the focus was on programs implemented by IIE but funded by various donors, including ECA. The researchers sampled ten different programs and carried out a survey of grantees who completed programs administered by IIE between 2005 and 2015 (Sanger, 2015). As their sample included some non-government funded exchange programs, this study, while relevant and a contribution in a space where much evidence is missing, still does not effectively address the gap in both the body of evidence and the literature around public diplomacy program evaluations.

Another related study reviewed thirty evaluations of existing and former programs completed in the past 15 years (2001–2016) seeking to identify effective ways to measure the effects of investments in international higher education scholarship funding worldwide (Martel, 2018). Across the thirty studies, the author found that the evaluations leveraged methodologies that illustrated change in the individual, which included categories such as completion and satisfaction, changes in academic and/or professional attributes, as well as change in personal attitudes or beliefs (Martel, 2018). Martel (2018) identifies some methodologies that might be useful in these types of programmatic evaluations: a counterfactual design contribution analysis, and social network analysis. For a counterfactual design, most often understood as a variation of a randomized control trials (RCTs), can definitively measure changes related to the program intervention but Martel identifies a variety of limitations in identifying a comparison group. One challenge is the inability to source a group that might be similar to international scholarship recipients – she proposes that a potential group that would serve effectively as a comparison group could be the semi-finalists that were equally qualified but did not ultimately receive the scholarship (Martel, 2018). A contribution analysis design, which would focus on the additive value of a program while still taking into account any external factors may have on the beneficiaries, builds a case for reasonably inferring causality to a reasonable extent considering the complexity of potential confounding factors (Martel, 2018). Social network analysis (SNA) as a methodology allows evaluators to identify impacts related to the network or connections for participants. Martel (2018) describes that using this methodology allows the evaluator to contextualize information about networks to understand how potentially effective networks are. The meta-review found that several programs were able to show that program networks lead to significant outcomes and new collaborations among its recipients. Martel (2018) identifies the Fulbright program specifically in framing an example of where SNA might have utility: "The Fulbright Program has an extensive program network that is comprised of current scholarship recipients and thousands of alumni. All these students and scholars together share a common characteristic (receipt of a Fulbright scholarship) and can be analyzed based on their network's properties" (p. 292).

Pahlavi also presents a critique of evaluating public diplomacy programming. Following a critical review of different studies focusing predominantly on the media arm of the public diplomacy mechanism, Pahlavi still makes some crucial observations and suggestions. He introduces both the strategy and limitation of using proxy indicators, which are short-term or immediate, to get at the measurement of harder to identify outcomes like the long-term impacts:

> For example, to evaluate PD's contribution to reinforcing mutual comprehension between the United States and the world, the indicator used is 'the percentage of participants who remain

> in contact with host country people met on their program one year or longer after their program'. The problem is that these replacement performance indicators do not address the ultimate outcomes expected for these programmes (Pahlavi, 2007, p. 258).

A proxy indicator is a strategy that enables evaluators to make a plausible correlation for an outcome that is challenging to capture data on. These proxies are considered signals that can evaluate progress towards longer term goals when an evaluator is only able to capture mid-term result data (Banks, 2011). An example might be the number of participants who gained access to new knowledge or activities that by participation would result in a change of understanding. In this example, there may be no opportunity to gather data to determine if that change of understanding did occur; however, a proxy indicator approach takes the audience number and infers a measurement that may correlate with the desired impact.

Summary and case for a coherent methodology

Despite efforts to measure public diplomacy writ large, additional context that was provided through a preliminary survey of the existing literature demonstrates that there seems to be large gap in knowledge regarding effective measurement of program-based people-to-people diplomacy programs. Though efforts have been made, there are still significant unknowns about the long-term impacts of U.S. government financial investment in this style of public diplomacy over the past 50 years. A report released by the Advisory Commission on Public Diplomacy (ACPD), a bipartisan committee established in 1948, underscored the reality that after over 70 years of formalized public diplomacy initiatives, despite different efforts, the impact of most of programs cannot be empirically verified (Rider, 2015). Pamment (2014) argues:

> The number of studies that consider the methods used for the evaluation of soft power and their theoretical grounds remain few compared to the number that discuss policy goals or campaign outputs…When describing how campaigns are evaluated, scholars tend to make assumptions based on the goals or outputs of an initiative rather than on the basis of reliable, empirical data on its results (p. 50).

Evaluation of federally-funded programming has grown in popularity since the 1960s and has been increasingly required through has through legislation or through committee and Member office requests

(Kaiser, 2010) in more recent years to support decision-making. Pamment (2014) claims that evaluation is not simply a function conducted according to a best practice, but instead represents a struggle over complex elements including meaning, power and knowledge within and between soft power organizations, their stakeholders, and their target publics (p. 51).

Carissa Gonzales (2015), after conducting dozens of personal interviews with public diplomacy leaders, concluded “Rigorous evaluation of our programs is crucial in order for interagency policymakers and Department leadership to see Public Diplomacy for what it actually is: an instrument of national power.” Another challenge, beyond the lack of cohesive evidence on different longstanding programs is that the evaluations that have been carried out often lack rigor (Rider, 2015). Rider (2015) critiques that public diplomacy evaluations have focused more on immediate outputs than longer-term outcomes and sometimes exaggerate results, and their function is ultimately an exercise in “placating Congress” (n.p.). The choices motivating evaluation practice for public diplomacy programming reflect complex power structures and often an idea is held by funders that leveraging best practices in evaluation for methodologies will lead to success (Pamment, 2014). Using the same methodology for all evaluation activities “implies both that PD campaigns function a particular way and that all soft power activities are conducted for the same reasons, which ignore complex structural and organizational concerns surrounding why and how PD activities are conducted, evaluated and justified” (Pamment, 2014, p. 51).

While a commonly affirmed conclusion is that measurement of impacts from public diplomacy programs are notoriously hard to measure as they involve intangibles that are considered soft outcomes. The challenge is then compounded by both the longer-term time frame in which such change can occur, as well as the confounding effects presented by participants’ contexts. What the literature does demonstrate is that there is clear evidence for the value of these exchanges in breaking down stereotypes, but there exists a distinct critique that exchanges should be structured to achieve maximum impact while being more diligent about measurement (Bellamy & Weinberg, 2008). Barriers to measuring impact are concluded to be as much organizational and operational as financial (Bellamy & Weinberg, 2008). Such barriers that stem from the challenges working across a dispersed community of federal agencies, NGOs, and academic institutions limit attempts to align efforts around a common set of objectives and standards for the evaluation of exchange programs (Bellamy & Weinberg, 2008). Banks, on behalf of the USC Center on Public Diplomacy, acknowledged the framing and provided insights into the gaps in evaluation of public diplomacy programs beyond methodology and in consideration of the larger field:

> The focus of most measurement effort has traditionally been on the PD "program," considered by many the primary unit of public diplomacy activity. We focus on the program… there are certainly benefits to focusing on "programs" as the central element in PD performance measurement. First, there is a long history of doing so, and many government-run PD programs have been around for decades. There may be, as a consequence, a fairly extensive, extant body of research—reports, surveys, program histories, and possibly even prior evaluations, not to mention usable baseline data—that can help provide background and context for new efforts to measure program success… each program is evaluated on its own merits, not comparatively… The downside of this approach—looking at each program in isolation—is that it may result in lost opportunities to see connections between programs, reach broader conclusions, and to provide more useful guidance to field practitioners. Integrated and/or crosscutting evaluations, while perhaps more time consuming, can also serve to increase contact among program offices, stimulate the sharing of ideas and best practices, and permit deeper insight into how to combine programs to best effect, with more efficient use of staff and financial resources. It may lead as well to consideration of other approaches to evaluation (Banks, 2011, pp. 25-26).

In acknowledgement of not only the lack of comprehensive evidence to date, as well as the push for decisions for public funding and policy to be driven by evaluative findings, there appears to be a cause, supported by the literature, to more strategically measure and evaluate public diplomacy-specific outcomes of the Peace Corps and cultural and educational exchange programs in order to determine global impacts that can better shape foreign policy moving forward.

Methodology

The methodology of this study was designed to test a few hypotheses regarding potential causes for the lack of a strategic, meta-approach to evaluation of public diplomacy programs funded by the U.S. government. This project seeks to test two specific hypotheses in its exploration: that evaluation has been affected by shifts in policy approaches and management internally, as well as the competition between organizations and the resulting lack of knowledge sharing.

Data collection and review

As this study sought to explore whether, prior to the launch of a codified indicator framework, the lack of standardization across programs within the ECA and across other agencies has limited strategic measurement *'mutual understanding*' of select U.S. government-funded programs that embody 'people-to-people' exchange. By establishing whether this limitation exists, it further sought to survey the different approaches to evaluation employed to determine impacts of such programming through conducting a meta-review, and) identify pain-points and opportunities for strategic measurement of 'mutual understanding' across diverse programs. This study relied on two sources: 1) a desk review of available, archival program documents/evaluations, literature on evaluation frameworks and public diplomacy; and 2) interviews with evaluation practitioners who are currently working on or have worked on evaluations for government-funded public diplomacy exchange programs. The data used in this study was collected through virtual or in person interviews with evaluators or sourced from open-access resources for program archival documents.

Interviews:

40 minute-long, in-depth semi-structured interviews were conducted with a sample of eight individuals that either identified as evaluators or individuals with evaluation as an aspect of their role who work on public diplomacy programming. For the purpose of this study, the term 'evaluator' will be used as descriptive of both types of professional. The study intended to interview those who identified as internal or external evaluator and have participated in evaluations of the seven sampled programs; however due to limitations the sample of interviews ranged from two program implementing individuals who had evaluation as part of their professional scope, two individuals internal to the State Department, one evaluator who had been contracted by ECA to perform an external evaluation, and three internal evaluators at two implementing partners of exchange programs. The interview protocol used for this study can be found in Appendix C:

Recruitment of evaluators into this research study drew from my personal network, as well as some identified through LinkedIn based on their organization or role. As this method for identifying individuals for interviews was not unbiased, I acknowledge significant limitations to this recruitment strategy as a researcher. However, I intentionally leveraged this strategy as it was intended to, and did, support one of my learning goals for this study: fostering collaborative, personal relationships with evaluators.

Desk Review

Information that was reviewed in the desk review for this study included:

- Program websites (including historical archives maintained by the programs)
- The U.S. Government Agency website that is the fiscal sponsor (ECA, Peace Corps)
- Implementing partner websites or materials/reports that were publicly available
- Any government laws/acts/reports relevant to the funding of these programs
- Promotional materials, infographics, or reports
- Available previously completed evaluations
- Peer-reviewed journal articles or publications related to evaluation, public diplomacy, foreign policy, etc.
- Existing evaluation/indicator frameworks within other sectors that were applicable

These materials were reviewed thoroughly and informed themes for the code tree to analyze both interviews and inform the meta-review of existing evaluation reports. A total of However, in trying to identify publicly available and relevant data sources, there were specific limitations that affected the study. The study had intended to include a review of evaluations from all seven identified programs; however, this was not possible due to challenges in accessing publicly facing reports. Some reports were not shared externally per a decision by the funding agencies, ECA or Peace Corps; others had never been conducted or were in the process of being conducted during the authoring of this study.

Data analysis

Based on the literature review, I created a code tree of themes that I anticipated would surface in the interviews and desk review. The code tree (see Appendix C on page 89) was updated iteratively through completing a thorough desk review of secondary data and archival program documents and key informant interviews.

Due to technological constraints, interviews were not recorded nor transcribed. During the interviews, I carefully took notes to summarize the individuals' responses. After completing the interviews, these notes were thematically analyzed for overarching themes that arise in evaluation of these programs based on the code tree. Sources were coded and analyzed thematically to identify cross-cutting themes, similarities and differences between the programs using the qualitative analysis software NVivo.

Limitations

This study was met with a variety of limitations. These included limitations related to extensive material available for the literature review but lacking for materials to support the desk review and analysis. Another limitation of the research design for interviews including a snowball sampling methodology drawing from the author's personal networks, which also led to challenges in confirming an interview with an individual who had worked on evaluation for the Peace Corps. A final, significant limitation was the author's own positionality in the field.

Material limitations for the desk review and meta-review analysis were largely due to issues of access or availability. Initially, the study intended to review a suite of evaluation reports as part of the meta-analysis, but there were few available. This was due either to issues of classification, effective record-keeping, or transferal of archival information to being published on public domains such as website. Additionally, due to ECA's approach to commissioning evaluations and budgetary limitations preventing more than a few evaluations to be commissioned and completed annually, several of the sampled programs did not have traditional evaluations completed at the time of this study. Additionally, contractual related privacy issues limited the ability of this study to take into account evaluation reports produced by implementing organizations and submitted to the State Department as they were not publicly available.

The snowball sampling used for identifying interview participants proved an asset in accomplishing one objective of this study but did introduce bias. Though strategic outreach was done through LinkedIn, cold contact emails, and through existing contacts, there were challenges in getting in contact with a variety of stakeholders to introduce different perspectives. For some of the evaluation firms that had previously been contracted by ECA to conduct an evaluation years ago, business mergers or individuals' professional transitions made it difficult to contact evaluators that had worked on those projects. While many individuals were approached for this study from a variety of relevant stakeholder groups, including ECA's Evaluation Division, Peace Corps, and implementing partners of public diplomacy programs including World Learning, Meridian, American Councils, IREX, Institute of International Education (IIE), Cultural Vistas, and FHI360, only a few individuals outside of the author's existing networks agreed to interview.

Finally, my position and role as a practitioner in this field both augmented and biased the findings of this study. Through a variety of personal interactions and professional activities, including working with the evaluation office at ECA or networking with evaluators from Peace Corps at conferences, as well as authoring sections in proposals that engaged with the materials used in this study, my positionality and background deeply influenced the research findings shared here.

Findings

Desk Research

In a review of available materials, including articles and various online resources, it became clear the shifting governmental policy and resources serve as a backdrop which has heavily influenced the challenge of measuring public diplomacy programs. Over the course of the nearly 80 years of educational and cultural exchanges, the desk review clearly revealed efforts to evaluate impacts of these programs have become progressively more important.

Summary of historical evaluation findings and efforts

Created in 1948, the U.S. Advisory Commission on Public Diplomacy (APCD) was created to appraise all public diplomacy efforts by the U.S. Government to report to the President, Secretary of

State, and Congress (U.S. Department of State, n.d.), including international broadcasting activities as well as exchanges. Starting in 1949, the APCD started publishing both annual and special reports (often multiple times a year) that began to furnish evidence for the impact of international exchanges through reporting to Congress (U.S. Department of State, n.d.). However, it wasn't until 1955, against the backdrop of the launch of the Fulbright Program and less than a decade after the passing of the Smith-Mundt Act, one of the first external evaluation studies was completed (Bettie, 2015). Bettie summarizes the findings of that study:

> Mendelsohn and Orenstein's 1955 study, the first to assess the Fulbright Program exclusively, included a post-sojourn questionnaire that considered effects on both the individual grantee and on his or her community upon returning home. They found that teacher grantees had organised international pen-pals for their students through contacts that they had made in the host country, supporting the Fulbright Program's aim of generating mutual understanding in both the home and host communities (Bettie, 2015, p. 35).

Wilson and Bonilla of International Research Associates, Inc. were also contracted by the Department of Sate to carry out multiple evaluation studies of exchange programs in 1955. Wilson and Bonilla (1955) acknowledged that formal evaluations of exchange programs utilizing techniques of social research were at that time a recent innovation. They also noted the emerging importance of evaluation in the field in referencing that in 1952, a committee on cross-cultural education had been appointed by the Social Science Research Council (SSRC) and had already begun a three-year program of research into the impact of American educational experience on students from other nations (Wilson & Bonilla, 1955). Wilson and Bonilla (1955) reviewed several of the evaluations they had conducted, including an evaluation in 1951 in Latin America and others in West Germany:

> These studies have concentrated solely on determining whether certain of the exchange programs have, in fact, succeeded in inducing change among individuals who, in turn, disseminate newly acquired information and ideas in their native lands… the not inconsiderable impact of foreign nationals on Americans here at home has been largely ignored, as have the reciprocal effects on Americans, and on the participating countries, of the visits of American students and specialists outside our shores. Only minimal use has been made of existing records and standard report forms, as well as of the rich but unorganized information available from administrators and others who have had years of intimate contact with exchangees; much remains to be done to systematize such record-keeping and observation in a way that would make them more useful to scientific evaluations (p. 30).

Including the above-mentioned reports by Mendelsohn & Orenstein and Wilson & Bonilla, the State Department commissioned a number of evaluations as a result of the growth of U.S.-sponsored exchanges, both academic and cultural, following World War II (Banks, 2011). These early efforts at evaluation, while creating a base of evidence, were not added to consistently or strategically by the State Department. Instead, evaluation remained under-resourced and sporadic for the next nearly half century, with anecdotes and old and/or limited studies normalized as the primary methods of recording programmatic successes.

In April 1963, the U.S. Advisory Commission on the International Educational and Cultural Affairs, authored a report called A Beacon of Hope: The Exchange-of-Persons programs. This report was commissioned on a special study of the effectiveness of exchange programs funded by DOS . The Government Accountability Office (GAO) published two reports in the late 1970s that looked at the infrastructure and existing practice in capturing impacts within the scope of their investigation. A 1979 report titled ‘Flexibility--Key to Administering Fulbright-Hays Exchange Program’ identified findings related to a study done after 10 years of the Fulbright program in Yugoslavia. This study included a questionnaire sent to 600 Yugoslavian Fulbright alumni, with around a 40% response rate (GAO, 1979, p. 49). The report summarized the findings:

> The responses showed that 83 percent thought their experiences were helpful, 51 percent advanced professionally as a result of the grant experience, 82 percent thought that their experience benefited their employer o/r organization, 85 percent maintain U.S. contacts, and 93 percent thought that educational exchange contributes to mutual understanding (GAO, 1979, pp. 49-50).

Also in the late 1970s, a dissertation identifies that there were at least 24 existing evaluation studies of the International Visitor Program (IVP) (Banks, 2011). Banks (2011) summarizes that Mueller, the author of a 1977 dissertation, found that the completed studies were inadequate, outdated and/or flawed, citing a lack of quantitative data and failure to identify any impacts of the program on host-country individuals.

Ad hoc evaluation efforts are referenced through the 1980s and 1990s, with evaluation being brought more to the forefront in the early 2000s. However, 1995, key legislation passed that would deeply affect the ability of government agencies to conduct evaluations: The Paperwork Reduction Act (PRA) (CDC, 2017). The PRA requires that agency must first seek clearance from OMB for review, which takes into account whether the proposed data collection is necessary; any changes to the proposed clearance package is then subject to being re-cleared if any changes (such as number of respondents or changes to the instruments) are made before being able to implement any of the revisions (CDC, 2017). This bureaucratic clearance process has now become part of most data collection efforts within American-based communities, which represents a significant number of public diplomacy program participants.

Fast forward to the early 21st Century, and paperwork trails and reports were much easier to locate. Snow (2009) summarizes an outcome assessment released by the U.S. State Department reporting very positive impacts from the Fulbright program. The assessment reported:

> …A nearly unanimous belief that participation in the Fulbright program promotes mutual understanding (97%) and builds leadership traits such as self-reliance and self-confidence (92%); ability to work closely with people from other cultures (89%); ability to lead others (71%); and willingness to lead others (70%). A full 100% developed a deeper understanding of their host country and 93% of the American students reported a deeper understanding of U.S. society and culture as a result of seeing it through the eyes of their host country citizens. Strong evidence for the multiplier effect was in evidence with two-thirds of the grantees maintaining collaborative ties with host country colleagues; nearly all encouraged friends and colleagues to participate in international exchange programs or to apply for a Fulbright grant. (Snow, 2009, p. 238).

The concept of evaluation began to be perceived as valuable by entities such as the U.S. government in the latter half of the 20th century, when appreciation of a more results-oriented approach to funding altered the paradigm in thinking about how donor-funded programs are managed, how practitioners interact with clients and volunteers, and how providers are held accountable. The State Department and independent agencies such as the Peace Corps were not immune to this paradigm shift and integrated evaluation into their structure through establishing evaluation divisions and creating program theories that framed the goals of exchange programs and started to define the outcomes that were expected as a result of participation in these programs.

Evaluation efforts at Peace Corps

Presently, Peace Corps centered evaluation activities in their 2014-2018 Strategic Plan, which sought to strengthen the agency's focus on evidence-based decision making, monitoring, and evaluation practices, as well as enhance the use of existing data and build a base of evidence (Peace Corps, 2018). As of 2010, the agency used only four primary tools to evaluate progress towards their three goals: the Volunteer Reporting Tool (VRF), the Project Status Report, Annual Volunteer Survey, and Results Based Field Evaluations also titled 'Impact Studies' (Peace Corps, 2010). The creation of the Office of Strategic Information, Research, and Planning to focus specifically on evaluation efforts was a significant step for the agency in the early 2000s (Peace Corps, 2009).

Following a comprehensive assessment completed in 2010, the Peace Corps adopted a strategy that has guided its operations and evaluation activities (Tarnoff, 2016). Peace Corps has only adopted a more standardized approach to evaluation across its many posts within the past ten years, including agency-wide standard indicators to allow reporting on common results across projects and countries (Tarnoff, 2016). Peace Corps currently utilizes a decentralized strategy for monitoring and evaluation that leverages overseas staff at posts to conduct country-specific programmatic monitoring and some evaluation activities, while larger-scale research and evaluation work occurs in a variety of headquarters offices, including the Office of Inspector General and the Office of Strategic Information, Research, and Planning within the agency (OSIRP) (Peace Corps, 2020). Each element is evaluated in a different way, though the overall goals across the Peace Corps are the same. For outcomes specific to the Volunteers' projects, a different objective is centered, and evaluation is done in exploration of those outcomes. Other types of categories that are evaluated also include the Peace Corps offices' abilities to train Volunteers effectively.

> For example, the objective of promoting sustainable change in the communities where volunteers work is measured by the percentage of projects with documented gains in community-based development outcomes. Underlying that indicator are efforts made in recent

> years to describe and document expected volunteer contributions to host community development goals. Another indicator of sustainable change performance will be the result of annual impact studies, an innovation launched in 2008 and used to develop best practices for agency programs. Other objectives are to enhance volunteer effectiveness (indicators include improved language learning, an improved site management system, and strengthened project planning); optimize volunteer resilience (indicators include increasing volunteer capacity to manage adjustment challenges and efforts to establish realistic expectations of service); build leaders for tomorrow (measured in part by the number of opportunities for RPCVs to engage in continued service) (Brown, 2019, p. 5).

The emphasis for activities that are evaluated are related to more output level findings: what happens to the Volunteers and how do they contribute while they are in their host country. Between 2008 and 2012, the Peace Corps conducted 24 impact studies on individual host country projects, which were conducted retrospectively by interviewing counterparts, beneficiaries, host country families, and stakeholders (Peace Corps, 2020). The topics of these studies topics include predominantly evaluation of programs (e.g., girls' education, youth camps, malaria prevention, HIV/AIDS awareness, and/or PCV contributions to sustainable change in host country programs) and operational topics (e.g., best practices in site development, working with local counterparts, and global and local partnerships; and/or mechanisms to engage communities in productive, collaborative, and inclusive relationships that foster development and strengthen cultural exchange) (Peace Corps, 2020). Evaluations related to the first two goals of Peace Corps, related to American PCVs abroad, focus more on the main program sectors and are often conducted by third-party research teams comprising of host country nationals (Peace Corps, 2010).

But for evaluating something like mutual understanding, efforts to capture longer term reciprocal impacts are needed. Efforts to accomplish this were outlined in Peace Corps 2009-2011 Strategic Plan but identifying the findings of these studies were proven to be challenging on the global scale. In the Strategic Plan, the report outlined that the agency would conduct field evaluations that include host country national perspectives of their understanding of Americans as a result of interacting with PCVs and look at the experience from various angles in order to show the knowledge, attitude, and behavior changes that occur within the host country populations through their work and contact with PCVs (Peace Corps, 2009). Efforts to accomplish this measurement was integrated into a survey approach by the Peace Corps OSIRP office, first piloted in 2015 and conducted again in 2016 (OSIRP), 2016), as well as through synthesis of different impact country studies (Rohrbaugh, 2016). In Rohrbaugh's 2016 study, she found that nearly two thirds of the total 928 counterparts, who were engaged across 24 different impact studies between 2009 and 2012, indicated that they had increased their understanding of Americans after working with Peace Corps Volunteers.

This comprehensive desk review was unable to obtain similar documentation for cross-cutting evaluative efforts to measure Peace Corps' achievement of fostering mutual understanding both prior to 2016.

Evaluation efforts at ECA

According to Ted Kniker (2011), the former Chief of Evaluation at ECA from 1998 to 2006, when he started, over 30 years after exchange programs had been cemented as part of U.S. public diplomacy strategy, "no documented literature about ECA evaluation existed, only internal reports from the 1980s when evaluation was limited to a handful of program reviews, generally conducted by academics or independent foreign-affairs specialists… Although the reviews discussed some program outcomes, the focus of these studies was primarily an assessment of whether the program was working as intended…. It was not evident that the evaluation reports were widely distributed or discussed, and so there was little programmatic investment into following up from the evaluations. Based on the number and scope of the reports that existed, one could draw the conclusion that evaluation was not considered essential to the bureau's operations" (p. 59). In the 1990s, Kniker recalls that more pressure surrounding budget justifications forced the leadership in ECA to integrate evaluation to provide data on how programs were performing so that resource allocations could be made. Bean (2015) argues that "as a result, the discourse of educational exchange has subtly shifted from one of mutual understanding, goodwill, and peace to one of 'impact,' 'effectiveness,' and 'accountability'" (p. 38). The ways that educational exchange contributes to the economic, political, and social goals of its primary funder—the U.S. federal government—have gained currency (Bean, 2015, p. 38). The pressure to identify contributions has led the Evaluation Division at ECA to expand and refine their scope for evaluation.

The ECA Evaluation Division is responsible for the ensuring that evaluation of their exchange programs occurs, but also highlights some specific goals for any evaluation findings:

- To inform strategic planning activities at both the Bureau and individual exchange program levels
- To generate evidence to determine whether an exchange program was implemented well, and if the program met its goals and objectives;
- To allow us to better understand both program successes and unintended consequences, as well as the ripple effect (how our exchange programs have impacted wider communities, both in the United States and throughout the world) (Evaluation Division (ECA), 2020).

In a report by the GAO (2003), the overall evaluation strategy for ECA was summarized:

> State's Bureau of Educational and Cultural Affairs surveys exchange program participants on their program experiences, their activities afterwards, and their impressions of the programs' effects on them. The bureau uses these and other data to evaluate specific exchange programs every 5 to 7 years on a rotating basis. The bureau has also recently initiated an effort to ask individuals who have completed exchange programs to recall specific attitudes and knowledge before the programs and how those had changed as a result of the programs. However, for most of its exchange programs, State does not systematically conduct pre- and post-program surveys that directly test and compare participant attitudes and knowledge before and after participation. Evaluation experts in the Bureau of Educational and Cultural Affairs acknowledged that conducting such surveys would provide more meaningful data on the effectiveness of exchange programs, but bureau officials estimated that such an approach would require approximately $2.2 million annually to pretest all alumni about their attitudes. It would also require two additional staff persons or hiring an evaluation firm to help with the data collection and analysis (p. 21).

Across the federal government under the George W. Bush administration, the Office of Management and Budget (OMB) recorded ways evaluation had started to be prioritized for ECA through the Program Assessment Rating Tool (PART), which was launched to rate and improve government programming. OMB in 2004 reported:

> ECA uses independent professional evaluators, selected through a competitive process, to assess the impact and outcomes of its programs. A multi-year program evaluation plan is developed every three years and adjusted as necessary to ensure that each year an academic, professional, and citizen exchange program is evaluated. All major funding line items have been evaluated at least once and will be re-evaluated every five to seven years. Where feasible, ECA has used comparison groups to more accurately assess the impact of the particular program. ECA also requires grantee organizations to submit evaluation plans for each grant. Some grantee orgs hire independent evaluators as well. While the evaluations focus on the "brand name" program activity, all evaluations are done in country-specific context. Recent evaluations include: Community Connections, Freedom Support Act Undergrad, English Language Programs, and MEPI Student Leaders… 21completed independent evaluations and 5 internal management reports of functional programs worldwide conducted by 15 external evaluation organizations have concluded that ECA's exchange programs are effectively

> meeting goals and administered well. ECA currently has four other evaluations in draft, 10 on-going projects, and four additional projects to be launched in 2004 (U.S. Office of Management and Budget, 2004).

Around a similar time frame in the mid-2000s, under Kniker's tenure as Chief of Evaluation, he oversaw the development of a first iteration of a bureau-wide attempt at measuring outcomes across programs by identifying key indicators and integrating them across evaluation projects. The Evaluation Division incorporated 10 required questions that covered key performance indicators into each evaluation project (Kniker, 2011). Through this approach, Kniker and his team were able to verify the performance data being collected and ultimately use that data to shape and focus future evaluative efforts. Kniker (2011) recalls "Eventually, we gained a high-level support to develop our own online performance measurement and evaluation tool, which automated our survey and reporting work" (Kniker, 2011, p. 63).

In a presentation shared in April 2010, the Evaluation Division reported that they had designed and implemented 54 outcome evaluations (either completed or on-going at that time). This presentation also referred to the Performance Measurement Initiative (PMI), which was designed to gather data on designated ECA outcome indicators (Evaluation Division of the Bureau of Educational and Cultural Affairs (ECA), 2010). Since 2004, in support of efforts across the federal government to monitor progress and improve performance the Evaluation Division administered surveys to provide feedback on the short-term outcomes of ECA exchange programs. An annual report presented by ECA's Evaluation Division from 2016 presented a summary of aggregated E-GOALS data (the Bureau's online performance management system) sourced from pre, post, and follow-up survey data across the ECA programs collected as part of the PMI portfolio in 2015 and 2016 (Evaluation Division of the Bureau of Educational and Cultural Affairs (ECA), 2010). The report drew on a combination of the standard indicator questions developed by Kniker and his team, which were deployed consistently across all programs to measure public diplomacy-related outcomes such as mutual understanding. The PMI measurement approach included eight main outcomes with corresponding indicators and was only integrated into a small portfolio of ECA program data collection. Yet while the PMI has since been discontinued, this model of identifying standard questions to streamline data collection was later included in the MODE framework approach.

A requirement of program contracts include evaluation, but ECA had not established any form of standardized approach as to how evaluations were to be done or what indicators should be measured for these types of programs until early 2020. The most recent effort to identify a strategy for measuring impacts of mutual understanding was undertaken by current Evaluation Chief, Natalie Donahue. In 2019, the ECA Evaluation Division led an initiative to redesign the performance monitoring process across the Bureau. In spring 2020, the evaluation division rolled out a new indicator framework called the 'MODE Framework'. The development of the Monitoring Data for ECA (MODE) Framework was a collaborative process in which the Evaluation Division worked closely with ECA program teams, senior leadership, award recipients, and regional Bureaus and select Embassies to create a tool that would ultimately be responsive to the Bureau's current data needs (Evaluation Division of the Bureau of Educational and Cultural Affairs (ECA), n.d.). The MODE Framework prioritizes monitoring data that leverages indicators designed to track program performance and support strengthened feedback mechanisms to promote more effective programs (Evaluation Division of the Bureau of Educational and Cultural Affairs (ECA), n.d.). The designated performance indicators have standard data collection questions to facilitate uniform data collection to ensure data validity and reliability (Evaluation Division of the Bureau of Educational and Cultural Affairs (ECA), n.d.). The MODE Framework is tied to current Bureau strategy as well as the National Security Strategy defined by the current administration (Evaluation Division of the Bureau of Educational and Cultural Affairs (ECA), n.d.). As the MODE Framework was recently launched, the Evaluation Division is still in the process of implementation for data collection efforts and has not yet proved whether it is supporting its intended goal.

Evaluation efforts of P2P by external parties

Doing a simple search for "evaluation of exchange programs" will generate a wide variety of articles and reports for evaluation efforts conducted by external parties (not government-hired contractors). There are numerous studies conducted by Fulbright alumni or RPCVs whose personal stake in the program has led them to deeper lines of inquiry, as well as other scholars. While an in-depth look at these types of reports was not a focus of this study, one notable difference of the types of reports emerged upon a cursory review: these reports often framed a critique of the program. One example was a study completed by Chinese scholars, Meirong Fu and Xin Zhao, where they observed that context specific aspects of exchange participants' experience goes often unacknowledged. In their comprehensive literature review, they identify that:

> Some studies did examine U.S. international exchange programs from other perspectives, notably the personal and professional effects on program participants and the ripple effects (Watkins, 1986; Dudden and Dynes, 1986; SRI International, 2005; ORC Macro, 2006; Scott-Smith, 2006; IAWG, 2013). Nevertheless, these studies largely make broad-based impressionistic evaluations with generalizations of the effects or document the positive outcomes (Fu & Zhao, 2017, p. 4).

As the pressure to justify funding and demonstrate positive outcomes related to programs for alumni and host communities does not apply to external scholars, they are able to explore the topic from a more neutral orientation that allows room for critique.

Challenges in sustaining evaluation efforts

Though evaluation has become an increasingly valued complement to programming, there are still a variety of challenges faced in attempting to measure the scope of impact for public diplomacy programming. These challenges include methodological feasibility, which can limit rigor of evaluation studies, longitudinal limitations caused by incomplete records, the global scale of participation and sometimes contradictions between programming and specific, discrete goals of U.S. foreign policy in different areas of the world, as well as shifting priorities of the government over time stemming from changes in administrations.

Methodological Challenges

This desk review, echoing the literature review, revealed that there have been on-going methodological challenges. During his tenure as Chief of Evaluation at ECA, Kniker (2011) observed:

> As with most evaluations, we struggled with issues around the purity of methodology and rigor versus feasibility within resources and environment. For example, we questioned how we could draw conclusions about programs that operate in 120 countries when we could only collect data in 5 or 6. Most of our programs drew participants through merit-based selection, and so we could not use randomized controlled trials as an experimental design, and had limited funds to conduct quasi-experimental designs. Appropriate comparison groups were difficult to construct within our target audiences, because of self-selection issues. We faced issues with questionnaire or interview delivery. In many countries, alumni were located outside of major cities, did not have reliable mail or Internet systems, and it was cost-prohibitive to send evaluators to them. (p. 64).

Kniker's quote highlights various methodological questions: sampling across programs with a global scale, as well as reaching target audiences to participate in surveys or interviews. What can be inferred by Kniker's mention of experimental design is that within public diplomacy program, it is near impossible to identify a comparison group to then attempt to identify specific contributions of the program experience for alumni. Across the desk review, there were mentions of issues with mitigating bias in a largely survey-based methodology. In establishing impacts on communities beyond individual participants, there was mention of this challenge even as early as the 1960s by the ACPD (U.S. Advisory Commission on International Educational and Cultural Affairs, 1963).

Longitudinal Limitations

Another reoccurring theme across the both the literature review and the desk review identified challenges related to longitudinal data collection. Zaharna (2009) identifies a potential cause for this challenge as being linked to the contracted time-frame of the program, which does not include long-term follow-up or mechanisms for expansion, and that the inherent design of exchange programs were not conceived or designed to be networking initiatives. Banks (2012) also noted that government record-keeping, specifically in countries where participants come from, is "spotty, so there may be no baseline data to start with, and little capacity to track cases over time" (p. 31-32). Kniker at ECA also calls this out:

> [A] challenge we faced was poor, or nonexistent, data. Because ECA programs were people-oriented programs, and our office was tasked with documenting the outcomes of these programs, contacting former participants was essential. We discovered there was not a central repository for alumni, and if information existed, it was kept by either the U.S. Embassies or by private, partner organizations. Much of the information was not up to date, except for

> American participants, who constituted about one-fourth of all alumni….Because data needs are usually not discussed in an enterprise-wide way, it leads many agencies to develop independent application systems and/or data storage… a successful evaluation had to draw upon data from the program office, the program partners, the grants coordination office, the budget office, and the U.S. embassies, yet none of these data pieces were captured in a single place… there is no alignment of the data, which can lead to duplication efforts, or data inconsistencies (Kniker, 2011, p. 63; p. 70).

Due to this spotty information, during Kniker's tenure, he reported that about a third of evaluation project budgets were allocated specifically into finding people and updating contact information to ensure that an evaluation reached a wide enough spread of alumni to capture some aspects of program impact (Kniker, 2011). This issue, in many ways, is compounded by the global scale of many public diplomacy programs. Though this effort is documented even by reports by the GAO, to date it appears it has not been effectively addressed. In a report from 2003, the GAO identifies that :

> The State would still have difficulty conducting long-range tracking of exchange participants because it lacks a database with comprehensive information on its various exchange program alumni. Although State's records are better for more recent exchange participants, its ability to locate individuals who participated prior to 1996 is limited. State had planned to begin building a new worldwide alumni database with comprehensive data, but Bureau of Educational and Cultural Affairs officials told us that State had received insufficient funds to do so (p. 21).

Directional Shifts in Federal Strategy and Foreign Policy

Always implicit in public diplomacy programming are United States' foreign policy aims. The ever-present tension with evaluating exchange programs is that the types of outcomes related to individual development or access to new skills or opportunities are much easier to measure than longer-term, complex outcomes intrinsic in public diplomacy. This is further complicated as each administration reauthors foreign policy priorities and federal procedures. Just looking at the past 20 years and the various administration shifts proves this point.

Under the George W. Bush administration, much of foreign policy was driven in reaction to the events of 9/11 and subsequent War on Terror. Decisions about what funding went where were informed according to specific goals aligned with the administration's priorities. Additionally, later in his tenure, President Bush launched a process for evaluating federal programs to ensure that the government was spending tax-payer dollars effectively: the PART, which was rolled out through an executive order in 2007 (Bush, 2007).

Under the Barack Obama administration, different procedures and priorities were established. President Obama sought to build on efforts of previous administrations and rolled out a new approach to inform strengthening the use of evidence in strategic and budgetary decisions (Chair of the Council of Economic Advisers, 2014). This new approach leveraged more impact evaluations, but also sought to embed evaluation into routine program operations and using existing data to support measurement. His foreign policy priorities also shifted and informed strategic changes in public diplomacy programming. One notable case is the creation of the Young African Leaders Initiative (YALI) and the Mandela Washington Fellowship program, a new public diplomacy program focusing on the African continent which is administered by ECA (Office of the Press Secretary, 2014).

Now under the Donald Trump administration, again new procedures and priorities have been established. Most notable are spending priorities to support American communities, which has led to ECA' adoption of a new Functional Bureau strategy through 2022 focusing on advancing four foreign policy goals:

ECA's operations through 2022 will chiefly be devoted to advancing four foreign policy goals for which international exchanges have a demonstrated positive impact, and to one management goal. ECA will:

1) Promote American leadership through people-to-people exchanges that advance American foreign policy objectives and deepen the mutual understanding that underpins U.S. relationships with foreign countries.

2) Renew America's competitive advantage for sustained economic growth by increasing the global skills of Americans and expanding the reach of U.S. businesses and institutions.

3) Counter foreign government disinformation and foster alternatives to radicalization through international exchange programs.

4) Promote American values through professional, educational, and cultural programs that bolster democratic principles and encourage strong civil society institutions, human rights, and independent media.

5) Ensure effectiveness of ECA programs and accountability to the American taxpayer by evaluating programs robustly, expanding use of virtual technologies, and leveraging relationships with program alumni (Bureau of Educational and Cultural Affairs, 2018, p. 4)

These types of policy pivots have repercussions on the span of grant-making for implementing partners, but also introduce ongoing challenges for evaluators.

Findings from meta-review of evaluative efforts

Profiles of sampled program evaluations

Snow (2009) argues that interpersonal communication is the most important opportunity to build trust, understanding, and friendship ("mutual understanding"). She further claims that this is why the Fulbright Program, International Visitors Leadership Program, and arts and writer exchanges have the most promise for winning hearts and minds in the United States (Snow, 2009, p. 5). The sample of seven noteworthy programs intended to be profiled in this paper included: the Fulbright Student Program, the Fulbright Scholar Program, International Visitors Leadership Program (IVLP), Professional Fellows Program (PFP), Youth Exchange and Study Program (YES), Future Leaders Exchange (FLEX) Program, Youth Ambassadors Program, and the Peace Corps. These programs were chosen because of the explicit nature of their overall goals as being inclusive of "mutual understanding". However, as there were no released reports to date for the Professional Fellows Program (PFP) or the Youth Ambassadors Program, these programs were removed from the sample. Please refer to Appendix A:Appendix A: for detailed overviews of programs included in this study, and Founded in 1961, the Peace Corps' three-point legislative mandate, unchanged since its founding, is to promote world peace and friendship by improving the lives of those they serve, help others understand American culture, and bring volunteers' experience back to Americans at home. To date, more than 235,000 Peace Corps volunteers have served in 141 countries and in September 2018, there were 7,367 volunteers serving in 61 nations. Peace Corps volunteers come from every U.S. state. The Peace Corps sends American volunteers to serve at the grassroots level in villages and towns across the world for a period of 24 months, which includes three months of technical and language training followed by two years of service. Volunteers support host communities in every region of the world, with assistance programs in agriculture, economic development, youth development, health (particularly HIV/AIDS programming), and education. Of its volunteers, 42% work in education, the largest programmatic sector, and 46% serve in sub-Saharan Africa, the largest region. Based on its activities, the Peace Corps is an agency of both international development and public diplomacy, and its efforts are to improve both the condition of poor communities overseas and other nations' perceptions of the United States (Brown, 2019).

Table D on page 83 for a simplified breakdown of these programs and their goals.

Of the seven programs that were selected for this study based on the criteria of a program goal for "mutual understanding", only two thirds of the ECA-funded programming had full evaluation reports available, while the remainder only published one-pagers of the findings or the executive summary from the full report. Peace Corps was the only program where multiple reports from different stakeholder groups were openly available, so there were four reports specific to Peace Corps that were included in this analysis while other programs were represented through only one program report or executive summary.

> Shelf-life refers to those evaluations that may be technically accurate, may have incredibly valuable data, but are communicated in a way that doesn't resonate with decision makers and end up sitting on shelves or being used as doorstops… One senior manager told me that if evaluations didn't "sing," meaning he couldn't get the main point, or at least become interested in continuing to read from an evaluation in 30 seconds or less, we had failed. Evaluation half-life, on the other end, were those reports that were written without any sensitivity to the political environment in which programs operate and thus were deemed "radioactive" by leaders. These are the reports that are never released for fear of the great damage it might cause the agency (Kniker, 2011, p. 70).

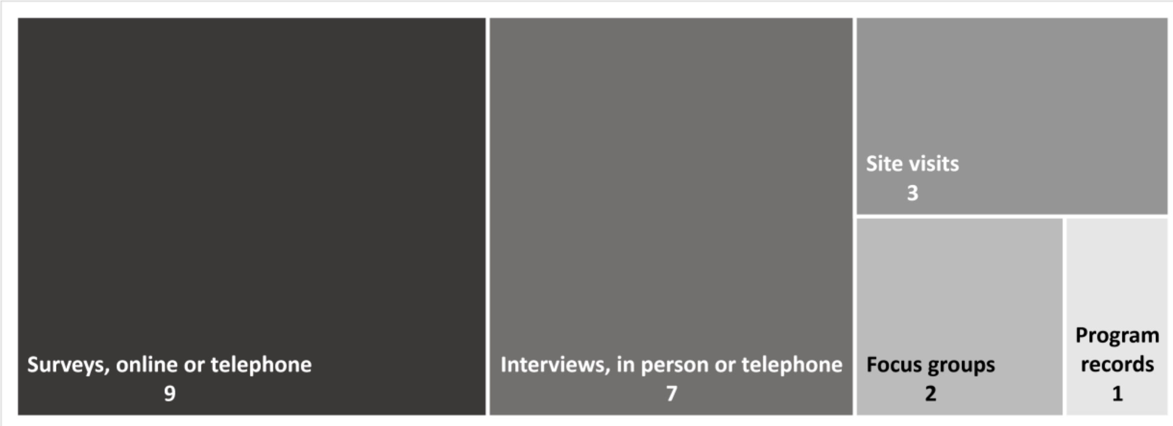
The evaluations that were conducted were done by either external contractors or by an office within the Peace Corps or State Department that is external to the program office. Some external contractors worked with subcontractors local to program participants' home countries to support with data collection. See Appendix A:Appendix B: for an overview of the evaluation studies included in this analysis.

Comparison of methodologies

Of the reports that were reviewed, the majority utilized a mixed methods approach for the evaluation. Though each of the studies identified that alumni had developed global relationships, interestingly this suite of reports did not leverage a methodology specific to measuring networks: Social Network Analysis (SNA). One study, for the FLEX program, did utilize an adapted experimental approach by including semi-finalists who did not ultimately participate in the program as a comparison group for the alumni.

The mixed methods almost always also relied on a survey, either administered in an online platform or via telephone, coupled with other qualitative methods. The below visualization outlines the different data collection strategies used across the different reports:

Figure 5. Data collection strategies used



Several of the qualitative approaches, specifically interviews, focus groups, or site visits largely took place in person during fieldwork conducted by the evaluation teams. However, for at least one report, an evaluation of Fulbright Foreign Student, virtual means were used to do interviews with alumni in one country. There was only one report that referred to conducting a review of program records to inform their study.

An additional aspect to methodological approaches included sampling. For each of these programs, one aspect of sampling was identifying specific cohorts of alumni to participate in the survey. A second step after specific program years were identified was identifying who from those cohorts would be invited to participate in data collection activities for these studies. For largely global programs such as the Peace Corps or Fulbright, there are often challenges in obtaining data and a tailored sampling strategy is often used. In the case of these reports, Peace Corps utilized a randomized sampling strategy across all of its hosting countries through outreach to individuals nominated by a PCV or to PCVs themselves. The reports evaluating the suite of Fulbright programs, on the other hand, leveraged a different sampling approach. One study of the Foreign Fulbright Student program sampled participants not based on country, but instead based on field of graduate study and based on participation in a specific program activity, the ‘From Lab to Market Seminar (FMLS)’. The evaluation of the Fulbright English Teaching Assistant (ETA) program’s sample was limited due to out-of-date alumni contact information, and only 77% of ETA alumni were contacted to complete the survey across all countries. sampled study participants for the qualitative data collection through focusing on specific countries which represented regional diversity, a range of program implementation models, and a high number of ETA participants. The sampling strategy of this study was also influenced by limitations due to combined fieldwork for two other evaluations of ECA language programs conducted by the same contractor. Another Foreign Fulbright Student study sampled participated only from 14 countries out of a global portfolio.

Several studies included efforts at triangulation of impacts through data collection with additional stakeholders, including the U.S. embassies, program staff, or beneficiaries such as alumni's employees or former students. The AWEP-IVLP evaluation methodology included interviews with individuals termed "impactees". These individuals were identified by the AWEP-IVLP alumni and were employees, business partners or community members who were impacted by the alumni following their completion of the program (General Dynamics Information Technology (GDIT); The District Communications Group (DCG), 2017). Across the Fulbright studies included in this study, data collection efforts at hosting institutions were also conducted among host colleagues.

Three reoccurring themes that came up across many reports included limitations related to effective data collection that can produce biased findings. These three main themes related to limitation s included 1) out-of-date alumni contact information, 2) inadequate records of beneficiaries, and 3) heavy reliance on retrospective, self-reported data. Across multiple reports, there appeared to be challenges in obtaining alumni contact information to ensure representation across the sampled cohorts. In two cases, the reports cited coordinated campaigns to update alumni contact information but both reports indicated that the full alumni cohort was unable to participate in data collection due to missing contact information. While these reports did not expand on whether the results of the survey may have been affected by these missing voices, there are implications as to whether the response rate achieved through the surveys was representative of the program alumni cohorts. For other studies, the evaluation team sought to triangulate findings reported from program participants by interviewing those who had either worked directly with the alumni during their exchange experience or those that had benefited from the alumni's experience since returning home Sampling strategies for these individuals often relies on the alumni to identify the potential respondent. This introduces bias as the alumni may not nominate individuals with the most comprehensive view of the impact of their exchange.

Synthesis of findings related to "mutual understanding"

As each of these programs had a specific programmatic goal of "mutual understanding", all of the studies explicitly looked at the achievement of this goal. The priority of these studies in assessing mutual understanding was less on identifying how much the alumni had learned about other countries, but mostly on how their understanding of the United States: the U.S. government and democracy, American diversity, and American values.

How mutual understanding was measured

For some reports, mutual understanding was measured through explicit questions. Other reports sought information about whether participants had a greater appreciation for democratic values. As the sampled program reports all had an explicit goal of mutual understanding, the line of inquiry of the evaluation questions did intentionally include trying to measure this phenomenon. Some examples for how researchers chose to approach this endeavor from the reports reviewed for this study are below.

For Peace Corps, they looked at mutual understanding as a contribution of interactions with Volunteers experienced by their host country counterparts. The research team's strategy for measurement of this was a MaxDiff scoring model: respondents ranked different categories as top or bottom choice, forcing prioritization and reducing priority bias. These rankings were then calculated into MaxDiff scores in aggregate, scaled from -100 to 100 (Peace Corps: Office of Strategic Information, Research and Planning (OSIP), 2016). They found that counterparts responded that they had learned about American values, how diverse Americans are in terms of race, religion or economic level, and about U.S. history and how the government works (Peace Corps: Office of Strategic Information, Research and Planning (OSIP), 2016). Across the comprehensive meta-review of Host Country Impact Studies also completed by the Peace Corps, they also sought to capture changes in understanding Americans (Rohrbaugh, 2016). In these different impact studies, researchers sought to measure change by asking counterparts to retrospectively rate their knowledge or understanding of Americans prior to meeting any PCVs, then asking them to rate their understanding after having worked with Volunteers (Rohrbaugh, 2016). To then measure whether exposure to PCVs resulted in a more positive attitude towards Americans, the researchers asked counterparts to reflect on what they thought of Americans after working with a PCV in a qualitative, open-ended format (Rohrbaugh, 2016). In a synthesis of this question across all studies, OSIRP identified that counterparts associated mostly positive traits to Americans, such as “kind”, “friendly”, “hardworking”, or “people like us” (Rohrbaugh, 2016, p. 36).

The IVLP study conducted by ORC Macro among alumni from Georgia, Kazakhstan, Russia and Ukraine framed their inquiry around mutual understanding in having alumni report on whether the program increased their understanding of the U.S. and American in general, U.S. values and culture, and the U.S. government. This study went one step further and asked a question explicitly about whether IVLP alumni agreed if the program was accomplishing its goal of increasing mutual understanding (ORC Macro, 2006). These were done largely as close-ended, quantitative questions and did not use a retrospective comparison approach. Another approach was to ask about the accuracy of information that alumni communicate about the U.S. and breaking down stereotypes in their home communities (ORC Macro, 2006).

In the YES evaluation completed by InterMedia, mutual understanding for participants was measured through asking YES participants to self-report in a survey if their stay in the U.S. increased their understanding of politics, government, economy, and the level of freedom and equality in the U.S. They also sought to capture this by identifying a measure of favorability. The report found that a large majority of alumni, a year after completing the program, rated that they had a “more favorable” view of the U.S. (InterMedia, 2009). Like the study done by ORC Macro, the researchers also sought to measure the multiplier effect and asked respondents to self-report on how they have informed friends, family and community members about the U.S. to breakdown stereotypes. The survey captured that respondents believed that these efforts had been successful and that people around them had a positive and nuanced view of the U.S. and of Americans through interacting with them (InterMedia, 2009).

The evaluation of the Fulbright English Teaching Assistant (ETA) program sought to measure mutual understanding by identifying what American participants brought into their foreign classrooms. The evaluators surveyed American ETA alumni asking if they brought their own [American] experiences, cultures and traditions into their classroom teaching (EurekaFacts, LLC, 2014). The survey also asked ETA alumni to self-report to what extent they broke stereotypes about Americans and if they changed people's impressions of the United States and Americans (EurekaFacts, LLC, 2014). Another piece to measure mutual understanding were questions related to whether ETAs had increased awareness of the U.S. government, rights and freedoms, or diversity in the country among their host students and colleagues. The mutual aspect of 'mutual understanding' in this report was addressed through survey questions asking ETA to self-report if they had increased their knowledge of their host country cultures and societies (EurekaFacts, LLC, 2014).

Are these programs accomplishing their goal?

According to these evaluations, the programs are indeed accomplishing their specific program goals as well as the broader, legislation-mandated goal of mutual understanding. What appears missing is any findings that might reflect negatively on the program. By and large, each report shared data that identified overall successes of the programs and painted a picture that the program had drastically altered the trajectory of alumni and oftentimes their close colleagues or family. While elements of this are likely true, the reliance on self-reported data and limitations in design to capture potentially less success-focused or nuanced findings limit the ability for evaluators to understand what specifically about participating in these programs accomplishes the goal of fostering mutual understanding.

Findings from the Interviews

In interviewing a diverse sample of evaluators with various levels of experience and different toles and expertise, some key findings emerged. The sample of evaluators included seven total evaluation professionals, with two thirds serving in primarily evaluation or research focused roles, while the other third had split roles supporting programming and evaluation functions. The experience of these individuals ranged from one and half years in an evaluation function and no advanced degree up to another evaluator with a PhD and around 15 years of evaluation-specific experience. Roles included a mix of individuals who identified as internal evaluators, either to a program implementing contractor, internal to ECA, or former external contractors. The breadth of experience and perspectives shared by this sample supplements the other findings of this study, illustrating a “lived experience” of making evaluation happen in public diplomacy programming.

Dynamics of evaluation contracts, questions, and program theory

For the evaluators that had experience as an external contractor, they shared that there were a variety of stakeholders that were consistently involved in their on-going work. For external contractors hired directly by the ECA evaluation division to conduct an impact evaluation, they worked very closely with a representative from the Evaluation Division and partnered with both the ECA grants officer and the implementing partner to ensure they had a good picture of what to evaluate. For representatives of the ECA evaluation division, they also had to work closely as the commissioning party and went through a lot of internal communication within ECA itself, as well as having to interact with other Bureaus at DOS and government offices (like the OMB) in order to support their evaluation efforts. For internal evaluators at implementing partners, they communicated frequently with the ECA grants officer and the implementing team to ensure that evaluation efforts were directed according to program needs and goals. One evaluator noted that since many public diplomacy programs have multiple stakeholders within the government at the planning and implementation level, like the program office or more than one federal agency, as well as regional bureaus & embassies, as well as the implementing partner, it's important to make sure that everyone is at the table the beginning.

A dynamic that revealed itself that is not uncommon to evaluation practice, but still notable, is the specific dynamic of funded contracts. As the commissioning partner, ECA holds a lot of decision-making power in terms of what is to be evaluated. This is either defined through 1) the request for proposals for a specific evaluation project or 2) the request for proposals for the program implementation grant. The specific goals and evaluation questions are then defined by the program officer in partnership with the evaluation division and are tailored towards specific strategic goals of the public diplomacy arm of the State Department. All evaluators interviewed observed that the funder's priorities came first in defining evaluation goals. For internal or external evaluators, this means that there is not much opportunity to be creative and explore other potential lines of inquiry as the agreed upon evaluation topics were laid out in a contract sometimes years in advance of the evaluation efforts taking place. For external contractors working directly with the ECA evaluation division, it seemed easier to identify potential other lines of inquiry, but for internal evaluators there were significant limitations. However, the limitations of inquiry are in many ways aligned with the larger bureaucratic goals of the DOS and ensuring evidence is available in how these programs are fulfilling their legislatively determined purposes. One evaluator observed that "when there's a change of administration, we have to shift how we talk about programs and what data we choose to collect." As evaluation has progressively supported more participatory engagement with program participants, the funder's priorities directly limit opportunities to implement innovative, participatory designs that seek to integrate program recipients understanding of 'success' and 'impact' into the research effort, as well as donor and program implementer concerns.

For evaluators who served an internal function, another dynamic that revealed itself was the ways of dissemination of findings. As internal evaluations are proposed in a contract, they are submitted but often not disseminated. The competitive context of non-profit implementers also directly contributes to the reporting and dissemination opportunities. One evaluator noted that evaluation efforts tend to be narrated through a more positive lens than what is objective out of concern for revealing any inadequacies of the contractor that might provide cause for the next contract to be awarded to another organization. This evaluator noted that “there was a looming presence of the recompete around the corner” in his efforts on an evaluation project and submitted report. All internal evaluators identified the same tension around having to measure only what was put in the proposal, which were sometimes not written by authors with much knowledge of evaluation. A federal evaluator observed: “People haven’t figured out a good way to use administrative data in evaluation or monitoring. There’s a lot of data that’s collected through participant information or reports, a wealth of information that hasn’t been synthesized or used yet.” Multiple interviewees also shared that, from their perspective working internally, there does not seem to be much clarity about use of evaluation efforts for public diplomacy programs at ECA or the Peace Corps as much of dissemination is related to accountability to Congress and that it seems public diplomacy evaluation efforts at large are trying to demonstrate success rather than learning.

Strategies and approaches for evaluating ‘mutual understanding’

All evaluators emphasized the importance of using mixed methods approaches in evaluating mutual understanding. One evaluator noted that this methodology usually includes a survey, interviews and focus groups and that it’s always hard to break away from that mold. One aspect of mixed methods that proved challenging were budgetary limitations that impacted time spent on analysis of more qualitative data. Conducting field work and obtaining data from interviews or focus groups is much more time intensive and cost-laden, so some internal evaluators noted that their level of mixed methods was sometimes just limited to open-text responses on a survey.

Three evaluators highlighted the importance of starting with a review of the literature to better understand hard-to-measure concepts like mutual understanding. Not only can a literature review reveal how this may have been approached previously but can also provide key contextual factors related to how definitions may change depending on the country and to support establishing a baseline understanding of what might define a concept of ‘mutual understanding’.

While only two evaluators shared that they based their evaluation strategy on the Kirkpatrick model, examples of competencies that evaluators shared that they sought to measure as components of mutual understanding included intercultural competence, gains in individual knowledge, impacts on the participant's identity, their sense of personal and ongoing connections to the international community, and an increased openness to diversity. In a more targeted approach, efforts to measure mutual understanding specifically interrogated participants' increased knowledge of the U.S. government, culture, values and relationships with Americans as a measure of increased *'mutual understanding.'*

Another strategy that was shared was the importance of trying to triangulate findings. Since much of evaluative efforts for public diplomacy programs rely on self-reported responses from program participants. One evaluator observed that it is hard to trust the validity of self-reported data as a large challenge is that there is often inflation with self-reported skill growth. Another evaluator observed that there was recall bias introduced by asking alumni to recall their experience years later. Another evaluator observed that since there are limited resources and budget and verifying self-reported data would potentially triple your budget, evaluators are somewhat forced to rely on what respondents say in surveys or interviews about their growth. The listed ways that this sample of professionals were working to triangulate data included interviews or focus groups with community members or surveying designated program-affiliated host contacts like supervisors or advisors.

Challenges associated with evaluating programs

As referenced above, a recurrent challenge mentioned by evaluators across all roles was related to budget. Budgets limited the ability to triangulate self-reported data, develop effect research instruments, conduct unrushed analyses, and to support dissemination of findings. The most immediate way that budget affected evaluation efforts for public diplomacy efforts was related to time available to spend. As much time, for both internal and external evaluators, is taken up with back and forth communication and collaboration with different project stakeholders, the remaining time to distribute the surveys, schedule and conduct interviews or focus groups, and then complete analysis and author a report is limited.

Additionally, due to the tendency to package evaluation findings as evidence of programmatic success, there is additional time taken up with back and forth in reviewing and finalizing the evaluation report.

Two listed challenges were limitations in triangulating self-reported data and updating participant contact information. Related to triangulation, for ECA's evaluation division, any commissioned evaluation that was targeting American citizens and residents was required to go through a clearance process with the OMB and any potential changes to a sample would have to be reapproved. Since this outreach requires a time-intensive clearance process, this effort is not made with evaluation efforts completed by implementing partners, so there are significant challenges for internal evaluators in gaining community perspectives. Site visits and sampling had to be done far in advance and sometimes did not reflect changes that happened in programmatic contexts. Another limitation was related to response rates – multiple evaluators commented on challenges in outreach to alumni populations to conduct evaluative efforts. For evaluators, time and budget in their contract was then used to support identifying updated contact information prior to dissemination of any data collection tools. Even for internal evaluators, with limitations in proposed scopes of work for implementing partners to have up-to-date contact information, they were often limited to lower response rates on surveys as a slice of the alumni demographic proved unreachable. This challenge impacts representation, as one evaluator observed that those who often represent more invisible groups, such as participants with disabilities or those from rural areas, which then means the respondent pool is missing key participant perspectives. One external evaluator described important contribution of the perspective of one particular alumnus perspective with a disability to the overall evaluation findings after conducting multiple focus groups across different site visits.

Discussion of Findings

In writing this capstone over the course of several years, the number of definitions I read for soft power and the role it plays in public diplomacy is numerous. However, the most notable pivots I observed in what affected public diplomacy evaluation efforts was a change in administration and the subsequent pivots in foreign policy goals. This personal observation was supported by both the interviews and the desk review. The desk review highlighted specific ways that different administration's leveraged evaluation (e.g. the Bush administration's PART) as well as the way their overall priorities influenced what ECA and Peace Corps sought to measure. Under the Trump administration, a large emphasis has been placed on understanding the impact of tax-payer funded programming on the U.S. and thus evaluation efforts have pivoted into focusing less of individual participants' gains or experiences and instead on what the host community gained or experienced. Several interviewed evaluators called out challenges related to this shift in priorities. One evaluator, in responding to requirements for evaluations, shared that "almost all evaluations have some questions around mutual understanding, foreign policy… but more and more we are starting to think about impact in the U.S. and the bi-directionality of exchanges." This also proves challenging as evaluation efforts are subject to political variability, as observed in an interview, which the results in new strategic priorities being applied to a program whose fundamental design has not changed, 70 years into administration.

Another aspect to evaluating public diplomacy programs that was revealed was the importance of flexibility. By and large, evaluation efforts for federally funded public diplomacy programs are inflexible, either in determined by request for proposals or by contractual agreements. By eliminating flexibility, especially when it comes to budgets, it directly impacted the methodology or approaches that could be used to determine impact of these programs. As evidenced in the methodologies of the sampled evaluation reports, as well as the interviews, it is obvious that there is a lot of weight placed on surveys in public diplomacy programs. If there was budget and time available, then the quantitative data collection could be complemented by qualitative interviews and focus groups but were subject to limitations. While different methodologies, such as Social Network Analysis (SNA) or participatory methodologies can support other interesting conclusions, the bread and butter of public diplomacy evaluation is proven to rely on some iteration of mixed methods. This was consistent across the evaluation reports and what was shared in interviews with evaluators. A quote from the literature by Dr. Katherine Brown (2017), captures this finding well:

> There should not be a one-size-fits-all method to understanding the impact of public diplomacy, as it will require a mix of them. And there must be a long-term commitment to collecting the various data points that reflect the complexity of public diplomacy work and the relationships they try to create and maintain. Simultaneously, leadership in any public diplomacy operation needs to communicate that impact evaluations are essential, while also encouraging risk-taking and an open, constructive way of analysing its results. (p. 122).

One evaluator from ECA shared in an interview that internally at ECA, they are looking at other methodologies, such as the success case method or SNA; however, this seems to be in the initial stages.

Another thing that came through strongly in the literature, the desk review, and the interviews, is there is a plethora of work that has been done working to evaluate these programs, though perhaps not consistently, but these things have not been shared or synthesized. An evaluator from ECA shared that in "the public diplomacy realm, people haven't figured out a good way to use administrative data in evaluation or monitoring. There's a lot of data that's collected through participant information or quarterly reports; a wealth of information that hasn't been synthesized or used yet." From the desk review, it appears some of this is due to challenges due to not having coherent, maintained program records, but from interviews it also revealed that there are a variety of reports required as part of the funding agreement for implementers, but these reports are often just "sit on desks, constantly chasing an approval" and that by the time they are approved, the reports are then out of date.

Ultimately most of these reports do not end up being shared, though many of them do portray the programs in a positive light. Across multiple interviews, evaluators shared that there is pressure to spin the findings in a way that is perceived positively, to defend the contract and prove the value to the funder. Even at ECA, the language of defense is used, though in reference to defending to Congress for sustained funding: "There's this difference with what will make the program look good if the program needs to be defended vs. what will make the program better". These findings align with what was revealed in the desk review that many program evaluations are not released unless it tells a certain story. This limits the efficacy of evaluation and creates an ongoing challenge to tell a story of success, rather than the reality of "what happened". Furthermore, if evaluations are always portraying what is going well, it limits the ability of evaluation to support improving the program.

Additionally, as evidenced in the literature, there are several different frameworks to structure public diplomacy program evaluation, but none of the evaluation reports and only two interviews referred to using a specific framework to structure their approach. This seems to be a gap between the reality of evaluation in practice and the academic study of public diplomacy – there is room to integrate these frameworks. Through ECA's efforts to standardize monitoring through the PMI and now the new MODE Framework, there is some structure that allows the Evaluation Division to aggregate data across all programs to identify progress towards goals around mutual understanding, but ultimately does not appear to be grounded in the more academic literature surrounding public diplomacy evaluation.

Reflection and Conclusion

As a practitioner who has, in the duration of time in writing this paper, has worked on evaluations of four public diplomacy programs, all funded by ECA, and is a RPCV, the personal and professional relevance of this study was comprehensive. It allowed me to reflect, not only on my experience as a participant, but on my experience as an evaluator seeking to better understand the infrastructure and underpinnings of my work. Part of why this took so long to write is that I learned very quickly that the field of evaluation is not static – it is constantly evolving and innovating and pushing the envelope forward to do better. I also learned that the global environment that public diplomacy is intended to influence is also constantly shifting as administrations domestically and abroad shift their focuses. As I personally wrestled with how to evaluate the Mandela Washington Fellowship for Young African Leaders (YALI), a leadership development program for 700 promising Africans with an alumni cohort of over 4,400 individuals, while balancing the requirements of our award contract, the limitations of our budget, and a shift in administration that no longer put emphasis on impacts happening from the work of these alumni in their communities but instead about what impacts happen in American communities from the six weeks they interacted with Americans in the U.S. Yet talking to other evaluators about strategies they use, perspectives they bring to the work, asking about their expertise and interests reinvigorated me and supported a push forward in my own work. Furthermore, understanding the detailed background, the history and different pressures experienced at ECA and Peace Corps, as well as at implementing organizations, helped me reflect on why my work is inclusive of what it is today.

As evaluation efforts for public diplomacy are continually evolving, this study indicates that there needs to be much more invested in understanding what already exists and synthesizing it into something useable. Some emerging recommendations from the findings of this study are:

- Foster buy-in for a cohesive, theory-grounded approach to public diplomacy evaluation for U.S. government-funded programs that is not uprooted upon a shift in administration. Furthermore, this theory-grounded approach should be informed by the variety of different frameworks available specific to evaluation of public diplomacy, not just leadership development programs. As the goals of public diplomacy largely transcend bipartisan politics and politicians on both sides of the aisle support these programs, it makes sense that there should be some static strategy or metrics that are maintained, specific to program design, rather than having to apply metrics specific to a particular administration's foreign policy. By enabling the ongoing measurement of the same metrics, it will also greatly aid in the ability to track longitudinal impacts of these programs over time. Both the desk review and interviews with evaluators revealed that there is a lot existing evidence that has not been synthesized or made external, either due to contractual limitations or fear related to sharing the findings.

- Identify a range of methodologies that can generate rigorous evidence but not necessarily relying on the same cut-and-paste methods would greatly add to the robustness of understanding impact for public diplomacy programs. If evaluators for public diplomacy programs were able to engage in more participatory methodologies or even implement learning feedback loops to iterate or improve programs based on the inputs from an evaluation, rather than only relying on evaluations as a method of compliance and reporting for Congress, the impact of U.S. public diplomacy could be captured at a new level. Especially when attempting to capture soft outcomes, such as attitudinal change, more qualitative, participatory methods could help frame how attitudes towards the U.S. and democracy change specific to cultural contexts and communities. Even if other methodological approaches are unavailable, leveraging alumni as an advisory board to support design of survey tools could greatly support capturing more nuanced and targeted data that would illuminate the question of mutual understanding.

- Advocate for these programs and their funding not on the basis of success, but on the basis of learning and global engagement. By eliminating the perception of fear linked to having to "defend" a

> particular program against being cut from a budget or defend your contract from being awarded to a different implementing partner, evaluation would be able to be leveraged to identify actual outcomes and impacts, instead of only trying to focus on positive ones. This would build the body of evidence around what specific experiences participants might have during these programs that lead to positive or negative outcomes, which profile of a participant is best suited (which would inform selection processes), and which types of programs (long-term, short-term, professional, academic, etc.) are making the biggest different for U.S. public diplomacy around the world.

This Independent Practitioner Inquiry Capstone (IPIC) sought to explore what factors have limited strategic measurement of outcomes of "mutual understanding" for U.S.-funded programs by defining both *'public diplomacy'* across the field of international relations, and *'mutual understanding'*, reviewing different approaches to evaluation used to determine impacts of such programming through and identifying pain-points and opportunities for strategic measurement of *'mutual understanding'* moving forward. The findings related to these factors included challenges related to shifting foreign policy priorities, fears around insecure funding, and challenges in measurement and limited methodologies. Through a thorough literature review, desk review and interviews with evaluation practitioners, this study was able to bring to light some opportunities for approaching public diplomacy evaluation in the future. However, much remains to be done to understand the implications of programs such as the Fulbright Student Program, the Fulbright Scholar Program, International Visitors Leadership Program (IVLP), Professional Fellows Program (PFP), Youth Exchange and Study Program (YES), Future Leaders Exchange (FLEX) Program, Youth Ambassadors Program, and the Peace Corps on global *'mutual understanding'* and the exercise of soft power.

References

Aguirre International. (2003). *Assessment of the Future Leaders Exchange (FLEX) Program.* Washington D.C.: Aguirre International.

Ailes, C., & Russell, S. (2002). *Outcome Assessment of the U.S. Fulbright Scholar Program: Executive Report.* Arlington: SRI International.

Armstrong, M. C. (2009). Operationalizing Public Diplomacy. In N. Snow, & M. P. Taylor, *Routledge handbook of public diplomacy.* New York: Routledge.

Atkinson, C. (2010). Does soft power matter? A comparative analysis of studnet exchange programs 1980-2006. *Foreign Policy Analysis, 6*(1), 1-22.

Banks, R. (2011). *A Resource Guide to Public Diplomacy Evaluation.* Los Angeles: Figueroa Press.

Bean, H. (2017). Strategic Communication and the Marketization of Educational Exchange. In J. Mathews-Aydinli (Ed.), *International Education Exchanges and Intercultural Understanding* (p. 79.96). Cham, Switzerland: Palgrave Macmillan.

Bell, S., Gillespie, M., & Wilding, C. (2016). Rethinking Models of Evaluation: Sustainability as. *International Sustainable Development Research Conference,* (pp. 1-25). Lisbon: Open Research Online.

Bellamy, C., & Weinberg, A. (2008). Educational and Cultural Exchanges to Restore America's Image. *Washington Quarterly, 31*(3), 55-68.

Bettie, M. (2015). Ambassadors unaware: the Fulbright Program and American public diplomacy. *Journal of Transatlantic Studies, 13*(4), 358-372. doi:10.1080/14794012.2015.1088326

Bhandari, R., & Belyavina, R. (2011). *Evaluating and Measuring the Impact of Citizen Diplomacy: Current Status and Future Directions.* New York: Institute of International Education.

Bliss, K. S. (2013). *A Framework for the Strategic Planning & Evaluation of Public Diplomacy.* Lisbon: NATO, Joint Analysis and Lessons Learned Centre.

Broadcasting Board of Governors (BBG). (2018). *BBG Strategic Plan 2018-2022: Information Matters: Impact and Agility in U.S. International Media.* Washington D.C.: Broadcasting Board of Governors (BBG).

Brown, K. (2017). Challenges in measuring public diplomacy. In J. McClory, *The Soft Power 30: A global ranking of soft power* (pp. 119-122). Portland.

Brown, N. (2019). *The Peace Corps: Overview and Issues (CRS Report No. RS21168).* Washington DC: Congressional Research Service. Retrieved from https://crsreports.congress.gov/product/pdf/RS/RS21168

Buhmann, A. (2017, January 3). *How to move the needle on moving the needle*. Retrieved from CPD Blog: https://www.uscpublicdiplomacy.org/blog/how-move-needle-moving-needle

Buhmann, A., & Sommerfeldt, E. (2019, December 19). *Understanding Practitioners' Perceptionsin PD evaluation*. Retrieved from CPD Blog: https://www.uscpublicdiplomacy.org/blog/understanding-practitioners%E2%80%99-perceptions-pd-evaluation

Bureau of Educational and Cultural Affairs (ECA). (n.d.). *About ECA*. Retrieved from United States Department of State: https://eca.state.gov/about-bureau

Bureau of Educational and Cultural Affairs. (2018). *Functional Bureau Strategy.* Washington D.C.: U.S. Government.

Bureau of Educational and Cultural Affairs. (n.d.). *International Visitor Leadership Program (IVLP): At-a-glance*. Retrieved from Bureau of Educational and Cultural Affairs: Exchange programs: https://exchanges.state.gov/non-us/program/international-visitor-leadership-program-ivlp

Bush, G. W. (2007, November 13). *Executive Order: Improving Government Program Performance*. Retrieved from The White House: President George W. Bush: https://georgewbush-whitehouse.archives.gov/news/releases/2007/11/20071113-9.html

CDC. (2017, October 11). *CDC Paperwork Reduction Act Compliance: Ensuring Quality and Reducing Public Information Collection Burdens*. Retrieved from Office of Science (OS): Center for Disease Control and Prevention (CDC): https://www.cdc.gov/os/integrity/reducepublicburden/index.htm

Chair of the Council of Economic Advisers. (2014). *Economic Report of the President.* District of Columbia: United States Government Printing Office.

Clinton, J. &. (2015, December). Looking to the Future. *Foreign Service Journal*, 36-41.

Coyne, J. (2018, September). How the Peace Corps Transformed the Foreign Service. Retrieved from http://americandiplomacy.web.unc.edu/2018/09/how-the-peace-corps-transformed-the-foreign-service/

Cull, N. J. (2009). Public Diplomacy before Gullion: The Evolution of a Phrase. In N. Snow, & M. P. Taylor, *Routledge handbook of public diplomacy.* New York: Routledge .

Department of State. (2017). *Department of State Program and Project Design, Monitoring, and Evaluation Policy.* Washington D.C.: Department of State. Retrieved from https://www.state.gov/wp-content/uploads/2018/12/Department-of-State-Program-and-Project-Design-Monitoring-and-Evaluation-Policy.pdf

Duffey, J. (2009). How Globalization Became U.S. Public Diplomacy at the End of the Cold War. In N. Snow, & M. P. Taylor, *Routledge handbook of public diplomacy.* New York: Routledge.

Eikenberry, A. M., & Kluver, J. D. (2004). The marketization of the nonprofit sector: civil society at risk? *Public administration review, 64*(2), 132-140.

Epstein, S., Lawson, M., & Gill, C. (2018). *Department of State, Foreign Operations, and Related Programs: FY2018 Budget and Appropriations (CRS Report No. R44890).* Washington D.C.: Congressional Research Service.

EurekaFacts, LLC. (2014). *Evaluation of ECA's English Language Programs: Fulbright English Teaching Assistant Program.* Rockville: EurekaFacts. LLC.

Evaluation Division of the Bureau of Educational and Cultural Affairs (ECA). (2010, April). Public Diplomacy Assessment: ECA Program Evaluation and Performance Measurement. Washington DC. Retrieved from https://eca.state.gov/files/bureau/eca_evaluation_assessing-public-diplomacy_apr2010.pdf

Evaluation Division of the Bureau of Educational and Cultural Affairs (ECA). (2020, May 31). *The ECA Evaluation Division*. Retrieved from United States Department of State: https://eca.state.gov/impact/eca-evaluation-division

Evaluation Division of the Bureau of Educational and Cultural Affairs (ECA). (n.d.). *Monitoring Data for ECA (MODE) Framework*. Retrieved 3 10, 2020, from United States Department of State: https://eca.state.gov/impact/eca-evaluation-division/monitoring-data-eca-mode-framework

Evaluation Division of the Bureau of Educational and Cultural Affairs (ECA). (n.d.). *Facts and Figures*. Retrieved from United States Department of State: https://eca.state.gov/impact/facts-and-figures

Fitzpatrick, K. R. (2009). *The Future of U.S. Public Diplomacy: An Uncertain Fate.* Brill.

Fitzpatrick, K. R. (2011). *U.S. Public Diplomacy in a Post-9/11 World: From Messaging to Mutuality.* Los Angeles: Figueroa Press.

FLEX. (n.d.). *About the program*. Retrieved from Discover FLEX: http://discoverflex.org/about-program/

Fu, M., & Zhao, X. (2017). Utilizing the Effects of the Fulbright Program in Contemporary China: Motivational Elements in Chinese Scholars' Post-Fulbright Life. *Cambridge Journal of China Studies, 12*(3), 1-26.

Fulbright Foreign Student Program. (n.d.). *About: Fulbright Foreign Student Program*. Retrieved from Fulbright Foreign Student Program: https://foreign.fulbrightonline.org/about/foreign-fulbright

Fulbright Scholar Program. (n.d.). *U.S. Scholar Program: Overview*. Retrieved from Fulbright Scholar Program: https://www.cies.org/program/fulbright-us-scholar-program

Fulbright U.S. Student Program. (n.d.). *About: English Teaching Assistant Programs*. Retrieved from Fulbright U.S. Student Program: https://us.fulbrightonline.org/about/types-of-awards/english-teaching-assistant-awards

Fulbright U.S. Student Program. (n.d.). *About: Overview*. Retrieved from Fulbright U.S. Student Program: https://us.fulbrightonline.org/about/

General Dynamics Information Technology (GDIT). (2017). *AWEP-IVLP Program Evaluation Report.* Washington D.C.: The District Communications Group (DCG).

General Dynamics Information Technology (GDIT); The District Communications Group (DCG). (2017). *AWEP-IVLP Program Evaluation Report.* Washington D.C.: Bureau of Educational and Cultural Affairs, United States Department of State.

Gonzales, C. (2015). The Evaluation Revolution in Public Diplomacy. *The Ambassadors Review*, 36-43.

Hart, J. (2013). *Empire of ideas: The origins of public diplomacy and the transformation of U.S. foreign policy.* New York: Oxford University Pree.

Hayden, C. (2012). *The rhetoric of soft power: Public diplomacy in global contexts.* Lexington Books.

Hoffa, W. W. (2007). *A History of US Study Abroad: Beginnings to 1965.* Lancaster, PA: Frontiers & The Forum on Education Abroad.

Hoffa, W. W., & DePaul, S. C. (2010). *A history of US study abroad: 1965-present.* Forum on Education Abroad.

ICF International. (2017). *Evaluation of the Fulbright Foreign Student Program: Impact on STEM Participants.* Rockville: ICF International.

InterMedia. (2009). *Evaluation of the Youth Exchange & Study Program: Final Report.* Washington D.C.: InterMedia.

Kaiser, F. M. (2010). Independent Evaluators of Federal Programs (CRS Report No. R41337). Retrieved from https://crsreports.congress.gov/product/pdf/R/R41337

Kelley, J. R. (2009). Between “Take-offs” and “Crash Landings”: Situational Aspects of Public Diplomacy. In N. Snow, & M. P. Taylor, *Routledge handbook of public diplomacy* (pp. 70-85). New York: Routledge.

Kennedy-Lugar Youth Exchange Program. (n.d.). *History*. Retrieved from Kennedy-Lugar Youth Exchange Program: https://www.yes-abroad.org/about/history

Kirkgoz, Y. (2017). Exploring the impact of the Fulbright Foreign Language Teaching Assistant program on grantees’ educational and cultural beliefs and practices. In J. Mathews-Aydinli (Ed.), *International Education Exchanges and Intercultural Understanding* (pp. 131-155). Cham, Switzerland: Palgrave Macmillan.

Kniker, T. (2011). Evaluation survivor: How to outwit, outplay, and outlast as an internal government evaluator. . *New Directions for Evaluation*, 57-72.

Martel, M. (2018). Tracing the spark that lights a flame: A review of methodologies to measure the outcomes of international scholarships. In J. e. Dassin, *International Scholarships in Higher Education* (pp. 281-304). Chan: Palmgrave Millan.

Matwiczak, K. (2010). *Public Diplomacy Model for the Assessment of Performance.* Austin: The University of Texas at Austin.

McClory, J. (2017). *The Soft Power 30–A Global Ranking of Soft Power 2017.* Portland.

Media Relations Office - University of Albany. (2011, September 8). *Higher Education in a Post-9/11 World: Q&A with UAlbany Assistant Professor of Educational Administration and Policy Studies Jason Lane.* Retrieved from NewsCenter, the World within Reach: University of Albany, State University of New York: https://www.albany.edu/news/16124.php

Milton C. Cummings, J. (2009, June 26). Cultural Diplomacy and the United States Government: A Survey. *Cultural Diplomacy Research Series*, pp. 1-15.

Morrow, R. A., & Torres, C. A. (2000). The State, Globalization, and Educational Policy. In N. C. Burbules, & C. A. Torres (Eds.), *Globalization and Education: Critical Perspectives* (pp. 27-56). New York: Routledge.

Mueller, S. (2009). The Nexus of U.S. Public Diplomacy and Citizen Diplomacy. In N. Snow, & M. P. Taylor, *Routledge handbook of public diplomacy.* New York: Routledge.

Nye, J. (2006, February). *Think Again: Soft Power.* Retrieved from Foreign Policy: https://foreignpolicy.com/2006/02/23/think-again-soft-power/

Nye, J. (2017). Soft power: the origins and political progress of a concept. *Palgrave Communications,* 1-3.

Office of Stategic Information, Research, and Planning: Peace Corps. (2016). *2016 Global Counterpart Survey Summary Report.* Washington D.C.: Peace Corps.

Office of the Press Secretary. (2014, July 28). *BACKGROUND & FACT SHEET: The President's Young African Leaders Initiative (YALI)*. Retrieved from The White House: President Barack Obama: https://obamawhitehouse.archives.gov/the-press-office/2014/07/28/background-fact-sheet-president-s-young-african-leaders-initiative-yali

ORC Macro. (2006). *International Visitor Leadership Program Outcome Assessment: Executive Summary.* Washington D.C.: ORC Macro.

Pahlavi, P. (2007). Evaluating Public Diplomacy Programmes. *The Hague Journal of Diplomacy*, 255-281.

Pamment, J. (2014). Articulating influence: Toward a research agenda for interpreting the evaluation of soft power, public diplomacy and nation brands. *Public Relations Review*, 50-59.

Patton, M. Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *American Journal of Evaluation, 22*(3), 329-336.

Peace Corps. (2009). *The Peace Corps Strategic Plan 2009-2014 and Performance Plan 2009-2011*. Washington DC: Peace Corps.

Peace Corps. (2010). *The Peace Corps: A comprehensive agency assessment.* Washington Dc: Peace Corps.

Peace Corps. (2018). *Strategic Plan 2018-2022 and Fiscal Year 2018-2019 Annual Performance Plan.* Washington DC: Peace Corps.

Peace Corps. (2020). *The Peace Corps' Congressional Budget Justification Fiscal Year 2020.* Washington DC: Peace Corps.

Peace Corps: Office of Strategic Information, Research and Planning (OSIP). (2016). *2016 Global Counterpart Survey Summary.* Washington DC: Peace Corps.

Productivity Commission. (2010). *Strengthening Evidence-Based policy in the Australian Federation, Volume 2: Background Paper.* Canberra: Productivity Commission.

Public Diplomacy Alumni Association. (2018). *About U.S. Public Diplomacy*. Retrieved from PublicDiplomacy.org: https://pdaa.publicdiplomacy.org/?page_id=6

Rider, J. (2015, December). Proving Public Diplomacy Programs Work. *Foreign Service Journal*, 19-22.

Rohrbaugh, K. G. (2016). *Peace Corps Works: A Cross-Sectional Analysis of 21 Host Country Impact Studies.* Washington DC: Peace Corps.

Sanger, J. (2019). *The Power of International Education: Impact Analysis of IIE-administered Programs, 2005-2015.* New York: Institute of International Education.

Scott-Smith, G. (2008). Mapping the undefinable: Some thoughts on the relevance of exchange programs within international relations theory. 616(1), 173-195. *The Annals of the American Academy of Political and Social Science, 1*(616), 173-195.

Scott-Smith, G. (2009). Exchange Programs and Public Diplomacy. In N. Snow, & M. P. Taylor, *Routledge handbook of public diplomacy* (pp. 50-56). New York: Routledge.

Scott-Smith, G. (2012, September). Still Exchanging? The History, Relevance, and Effect of International Exchange Programs. Retrieved from https://www.e-ir.info/2012/09/14/still-exchanging-the-history-relevance-and-effect-of-international-exchange-programs/

Sevin, E. (2015). Pathways of connection: An analytical approach to the impacts of public diplomacy. *Public Relations Review*, 562-568.

Sevin, E. (2017). A multilayered approach to public diplomacy evaluation: Pathways of connection. *Politics & Policy*, *45*(5), 879-901.

Snow, N. (2009). Rethinking Public Diplomacy. In N. Snow, & M. P. Taylor, *Routledge handbook of public diplomacy* (pp. 3-11). New York: Routledge.

Snow, N. (2009). Valuing Exchange of Persons in Public Diplomacy. In N. Snow, & M. P. Taylor, *Routledge handbook of public diplomacy.* New York: Routledge.

Snow, N., & Taylor, M. P. (2009). Introduction. In N. Snow, & M. P. Taylor, *Routledge handbook of public diplomacy* (pp. ix-xi). New York: Routledge.

SRI International. (2005). *Outcome Assessment of the Visiting Fulbright Student Program: Executive Summary.* Arlington: SRI International.

Steven, D. (2007). Evaluation and the new public diplomacy. *Presentation to the 842nd Wilton Park Conference: The Future of Public Diplomacy* (pp. 1-20). River Path Associates.

Stromquist, N. P., & Monkman, K. (2014). Defining Globalization and assessing its implications for knowledge and education, revisited. In N. P. Stromquist, & K. Monkman (Eds.), *Globalization and education: Integration and contestation across cultures* (pp. 1-19). Lanham, MD: Rowman & Littlefield.

Tarnoff, C. (2016). *The Peace Corps: Current Issues (CRS Report No. RS21168).* Washington DC: Congressional Research Service.

Taylor, P. M. (2009). Public Diplomacy and Strategic Communications. In N. Snow, & M. P. Taylor, *Routledge handbook of public diplomacy* (pp. 12-16). New York: Routledge.

Thomas, I. (2018, January 19). *British Council on evaluating arts & soft powerprogramming.* Retrieved from CPD Blog: https://www.uscpublicdiplomacy.org/blog/british-council-evaluating-arts-soft-power-programming

Tikly, L. (2004). Education and the new imperialism. *Comparative Education*, *40*(2), 173-198.

U.S. Advisory Commission on International Educational and Cultural Affairs. (1963). *A beacon of hope: The exchange-of-persons program.* Washington D.C.: United States Government.

U.S. Department of State. (2005). *Cultural Diplomacy: The Linchpin of Public Diplomacy". Report of the Advisory Committee on Cultural Diplomacy.* Washington D.C.: U.S. Department of State.

U.S. Department of State. (n.d.). *Reports - U.S. Advisory Commission on Public Diplomacy*. Retrieved May 10, 2020, from U.S. Department of State: https://www.state.gov/reports-u-s-advisory-commission-on-public-diplomacy/

U.S. Department of State. (n.d.). *U.S. Advisory Commission on Public Diplomacy*. Retrieved May 10, 2020, from U.S. Department of State: https://www.state.gov/bureaus-offices/under-secretary-for-public-diplomacy-and-public-affairs/united-states-advisory-commission-on-public-diplomacy/

U.S. General Accounting Office (GAO). (2003). *U.S. Public Diplomacy: State Department Expands Efforts but Faces Significant Challenges (GAO-03-951).* Washington D.C.: U.S. Government.

U.S. Office of Management and Budget. (2004). *Detailed information on the global educational and cultural exchanges assessment*. Retrieved June 1, 2020, from ExpectMore.gov: https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/omb/expectmore/detail/10002214.2004.html

Vanc, A. M. (2016). Scope and status of public diplomacy research by public relations scholars, 1990–2014. *Public Relations Review, 42*(3), 432-440.

Watson, T. a. (2007). *Evaluating Public Relations: A Best Practice Guide to Public Relations Planning, Research and Evaluation.* Philadelphia: Kogan Page Limited.

Weerawardena, J., McDonald, R. E., & Mort, G. S. (2010). Sustainability of nonprofit organizations: An empirical investigation. *Journal of World Business, 45*, 346-356.

Wilson, E. C., & Bonilla, F. (1955). Evaluating exchange of persons programs. *Public Opinion Quarterly*, *19*(1), 20-30.

Wilson, I. (2014). *International Education Programs and Political Influence: Manufacturing Sympathy?* New York: Palgrave MacMillian.

Wyne, A. S. (2009). Public Opinion and Power. In N. Snow, & M. P. Taylor, *Routledge handbook of public diplomacy*. New York: Routledge.

Zaharna, R. (2009). Mapping out a Spectrum of Public Diplomacy: Information and Relational Communication Frameworks. In N. Snow, & M. P. Taylor, *Routledge handbook of public diplomacy*. New York: Routledge.

Zimmerman, R. (2015, December). Soft Power, High Impact. *Foreign Service Journal*, 23-24.

Appendices

Appendix A: Sampled Program Backgrounds

Fulbright Program

The Fulbright Program is an umbrella for a wide variety of biexchanges, both in bringing foreign nationals to the United States as well as sending American abroad. The program is considered a flagship as it originated with legislation in the 1940s. It is considered a prestigious award, and these individual grants comprise of different monetary amounts for differing durations of time, depending on the program. However, the original design of the Fulbright program was “to increase mutual understanding between the people of the United States and the people of other countries by means of educational and cultural exchange; to strengthen the ties which unite us with other nations by demonstrating the educational and cultural interests, developments and achievements of the people of the United States and other nations and the contributions being made toward a peaceful and more fruitful life for people throughout the world; to promote international cooperation for educational and cultural advancement; and thus to assist in the development of friendly, sympathetic, and peaceful relations between the United States and other countries of the world” (Snow, 2009, p. 7). Their website claims that more than 360,000 Fulbrighters from the United States and other countries have participated in the Program since its inception in 1946. Currently, the Fulbright Program operates in over 160 countries worldwide (Fulbright U.S. Student Program, n.d.).

Programs that fall under the Fulbright Program Portfolio included in this study are:

U.S. Student Fulbright Program

This program is for U.S. citizens who are graduating college seniors, graduate students, young professionals and artists to study, conduct research, and/or teach English abroad. There are special programs in addition to these two general types that include the Fulbright-Fogarty Award, as the

Fulbright-National Geographic Digital Storytelling Fellowship Award, as well as the Critical Language Enhancement award. The Fulbright U.S. Student Program offers research, study and teaching opportunities in over 140 countries to recent graduates and graduate students. These grant award lengths and dates vary by country, with the shortest being around 6 months and the longest being around a year. Approximately 1,900 grants are awarded annually to U.S. citizens in all fields of study. As written on the program's website, the Fulbright U.S. Student Program is the largest U.S. exchange program offering opportunities for students and young professionals to undertake international graduate study, advanced research, university teaching, and primary and secondary school teaching around the globe (Fulbright U.S. Student Program, n.d.)

English Language Teaching Assistant (ETA) Program

Falling under the range of offerings for Fulbright U.S. Student is the specific ETA program. As written on the program website, this program places Fulbrighters in classrooms abroad to aid local English teachers in classrooms ranging from kindergarten to university level. ETAs then help teach English language while also serving as cultural ambassadors for the U.S. (Fulbright U.S. Student Program, n.d.).

Foreign Student Fulbright Program

This program enables graduate students, young professionals and artists from abroad to study and conduct research in the United States. These Fulbrighters qualify for J1 visas and are able to enroll in formal education at universities or other educational institutions. Another component of this program is the Foreign Language Teaching Assistant (FLTA) track. The Fulbright FLTA Program supports teaching assistantships in over 30 languages at hundreds of U.S. institutions of higher education and is designed to develop Americans' knowledge of foreign cultures and languages by learning from these foreign nationals. Grants vary by country of origin and specific program but can range from up to 9 months to multiple years in the United States. According to their website, Fulbright Foreign Students can come to the U.S. from than 155 countries worldwide. Approximately 4,000 foreign students receive Fulbright scholarships each year (Fulbright Foreign Student Program, n.d.)

U.S. Scholar Fulbright Program

As written on the program website, the Fulbright U.S. Scholar Program offers approximately 470 teaching, research or combination teaching/research awards in over 125 countries per year. Opportunities are available for college and university faculty and administrators as well as for professionals, artists, journalists, scientists, lawyers, independent scholars and many others (Fulbright Scholar Program, n.d.).

International Visitors Leadership Program (IVLP)

IVLP is referred to as the U.S. Department of State’s premier professional exchange program. According to the ECA website, through short-term visits to the United States, current and emerging foreign leaders in a variety of fields participate in a series of professional meetings which reflect the participants’ professional interests and allow them to experience the U.S. firsthand and cultivate lasting relationships with their American counterparts. This program is typically only a few weeks long and according to the website, an estimated 5,000 International Visitors come to the U.S. each year. Since the program started in 1940, more than 200,000 International Visitors have engaged with Americans through the IVLP, including more than 500 current or former Chiefs of State or Heads of Government (Bureau of Educational and Cultural Affairs, n.d.). Approximately 5,000 foreign nationals visit the United States annually through the IVLP. Around 345 former and current heads of government have visited through participating in the program, including individuals such as Margaret Thatcher, Tony Blair, Indira Gandhi, Nicholas Sarkozy and Nobel Laureate Oscar Arias, all of whom participated early in their careers. In fact, two current Latin American presidents, Brazil’s Dilma Roussef and Uruguay’s Tabare Vazquez, are also among the 200,000 foreign alumni from 190 countries who have taken part in the program over the past 75 years (Zimmerman, 2015, p. 23).

Youth Exchange and Study Program (YES)

According to the program website, the YES program was established by Congress in October 2002 in response to the events of September 11, 2001 and provides scholarships for high school students from countries with significant Muslim populations to spend up to one academic year in the United States. During the program, students live with host families, attend high schools, engage in activities to learn about American society and values, acquire leadership skills, and help educate Americans about their countries and cultures. As of 2009, the counterpart, the YES Abroad program, was established in order to provide a similar experience for U.S. students (15-18 years) to spend an academic year in select YES countries (Kennedy-Lugar Youth Exchange Program, n.d.).

Future Leaders Exchange (FLEX) Program

According to the program website, FLEX provides merit-based scholarships for students from secondary school students from Armenia, Azerbaijan, the Czech Republic, Estonia, Georgia, Greece, Hungary, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Mongolia, Montenegro, Poland, Romania, Serbia, Slovakia, Tajikistan, Turkmenistan, and Ukraine to travel to the United States, live with a host family, and attend a U.S. high school for a full academic year. The program was established in 1992 and over 26,300 students have participated in the FLEX program to date. The primary goal of the FLEX program is to improve mutual understanding and develop and strengthen long-term relationships between citizens of the United States and other peoples and countries and was created with support from former Senator Bill Bradley. Bradley believed that the best way to ensure long-lasting peace and understanding between the U.S. and the countries of Eurasia is to enable young people to learn about the U.S. and Americans firsthand, and to teach Americans about their countries. The program ultimately served as the model for the YES (Kennedy-Lugar Youth Exchange and Study) and A-SMYLE (American Serbia and Montenegro Youth Leadership Exchange) programs and has since expanded to other countries (FLEX, n.d.)

Peace Corps

Founded in 1961, the Peace Corps' three-point legislative mandate, unchanged since its founding, is to promote world peace and friendship by improving the lives of those they serve, help others understand American culture, and bring volunteers' experience back to Americans at home. To date, more than 235,000 Peace Corps volunteers have served in 141 countries and in September 2018, there were 7,367 volunteers serving in 61 nations. Peace Corps volunteers come from every U.S. state. The Peace Corps sends American volunteers to serve at the grassroots level in villages and towns across the world for a period of 24 months, which includes three months of technical and language training followed by two years of service. Volunteers support host communities in every region of the world, with assistance programs in agriculture, economic development, youth development, health (particularly HIV/AIDS programming), and education. Of its volunteers, 42% work in education, the largest programmatic sector, and 46% serve in sub-Saharan Africa, the largest region. Based on its activities, the Peace Corps is an agency of both international development and public diplomacy, and its efforts are to improve both the condition of poor communities overseas and other nations' perceptions of the United States (Brown, 2019).

Table D

PROGRAM NAME	FUNDING AGENCY /IMPLEMENTING CONTRACTOR	DESCRIPTION OF PROGRAM	DURATION OF PROGRAM	REFERENCE TO "MUTUAL UNDERSTANDING"
Fulbright Student Fulbright Scholar	ECA IIE; IREX; CIES; Fulbright Commissions	The Fulbright Program is the flagship international educational exchange program sponsored by the U.S. government and is designed to **increase mutual understanding** between the people of the United States and the people of other countries. The Fulbright Program operates in more than 160 countries worldwide and has provided approximately 370,000 participants with the opportunity to study, teach, or conduct research in each other's countries and exchange ideas. Approximately 8,000 competitive, merit-based grants are awarded annually in most academic disciplines and fields of study.*	Many different types of Fulbright programs with different durations.	Program description uses "mutual understanding" explicitly as a goal. See highlighted text in Program Description.
International Visitors Leadership Program (IVLP)	ECA IIE; Global Ties; Cultural Vistas; Meridian; FHI360	Launched in 1940, the IVLP helps strengthen U.S. engagement with countries around the world and cultivate lasting relationships by connecting current and emerging foreign leaders with their American counterparts through short-term visits to the United States. The majority of IVLP exchanges include visits to four U.S. communities over three weeks, although projects vary based on themes, Embassy requests and other factors. Participants meet with professional counterparts, visit U.S. public and private sector organizations related to the project theme and participate in cultural and social activities.* Launched in 1940, IVLP seeks to **build mutual understanding** between the U.S. and other nations through short-term visits to the U.S. for current and emerging foreign leaders. Each year nearly 5,000 International Visitors come to the U.S. on the IVLP. More than 200,000 International Visitors have engaged with Americans through the IVLP, including over 335 current or former chiefs of State or heads of government. (https://eca.state.gov/about-bureau-0/organizational-structure/office-international-visitors)*	3-5 weeks	Program description highlights lasting "relationships" rather than mutual understanding. However, on the State Department website relating to the Office of International Visitors, the implementing arm of the program, "mutual understanding" is used explicitly. See highlighted text.
Youth Exchange and Study Program (YES)	ECA AFS; AMIDEAST; World Learning; others	**A)** YES Abroad program provides merit-based scholarships for eligible high school students to develop a perspective of a Muslim culture first-hand. Participants spend an academic year attending a high school while living with a host family in select countries with significant Muslim populations. A network of support through partner international exchange organizations, field staff, trained volunteers, and carefully selected host families helps ensure a successful exchange experience. Participants serve as "youth ambassadors" of the United States, **promoting mutual understanding** by forming lasting relationships with their host families and communities.* **B)** The YES program provides scholarships for high school students (ages 15-17 years) from countries with significant Muslim populations to spend up to one	Academic year 9-10 months	One side of the YES Program description uses "mutual understanding" explicitly as a goal (Americans going overseas). See highlighted text in Program Description. Necessary to highlight that for YES Program (Foreigners coming to the U.S.), this goal is

		academic year in the United States. Students live with host families, attend high school, engage in activities to learn about American society and values, acquire leadership skills, and educate Americans about their countries and cultures.*		not expressed explicitly though is inherent in the program itself.
Future Leaders Exchange (FLEX) Program	ECA American Councils; others	The Future Leaders Exchange (FLEX) program is a competitive, merit-based scholarship program funded by the U.S. Department of State. FLEX students spend an academic year in the United States living with a volunteer host family and attending a U.S. high school. FLEX students gain leadership skills, learn about American society and values, and teach Americans about FLEX countries and cultures. The primary goal of the FLEX program is to **improve mutual understanding** and develop and strengthen long-term relationships between citizens of the United States and other peoples and countries. There are currently 17 countries that participate in the FLEX program. These countries include Armenia, Azerbaijan, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Mongolia, Montenegro, Poland, Romania, Serbia, Tajikistan, Turkmenistan and Ukraine.*	Academic year 9-10 months	Program description uses "mutual understanding" explicitly as a goal. See highlighted text in Program Description.
Peace Corps	Peace Corps	The Peace Corps is a service opportunity for motivated changemakers to immerse themselves in a community abroad, working side by side with local leaders to tackle the most pressing challenges of our generation. The Peace Corps Mission: **To promote world peace and friendship by fulfilling three goals:** 1) To help the people of interested countries in meeting their need for trained men and women. 2) To help promote a better understanding of Americans on the part of the peoples served. 3) To help promote a better understanding of other peoples on the part of Americans.***	27 months	In talking about PC history, "mutual understanding" is used (https://www.jfklibrary.org/JFK/JFK-in-History/Peace-Corps.aspx). Peace Corps does not explicitly use mutual understanding, but Goal 2 & 3 are reciprocal in establishing understanding.

*Program description was taken from ECA website
**Program description was taken from Program website
***Program description taken from Peace Corps website

Appendix B: Evaluation Studies Summary

Program	Year of Evaluation	Evaluator	Target Sample	Regional	Methodology
ACADEMIC EXCHANGE					
Future Leaders Exchange (FLEX)	2002-2003	Aguirre International; University of Iowa Social Science Institute supported by local research organizations	Participants from Russia, Ukraine, Armenia, and Uzbekistan between 1993-2000	Russia, Ukraine, Armenia, Uzbekistan	Mixed methods: Survey, interviews, focus groups
Youth & Exchange Study Program (YES)	2003-2009	InterMedia	1856 participants across 4 cohorts between 2003-2009, secondary school students aged 15-17 from countries with a large Muslim-population	Global, 26 countries across MENA, South Asia, West Africa, & Southeast Asia	Mixed methods: Online surveys, focus groups
Foreign Fulbright Student *(From Lab to Market Seminar)*	2013-2017	ICF International	STEM-focused Foreign Fulbright alumni from 2007-2011 who participated in four-day enrichment seminar, From Lab to Market (FMLS)	Global; site visits in Brazil, Colombia, and Indonesia; virtual interviews done in Pakistan	Mixed methods: Online Survey, Interviews, Observation during site visits
Visiting Fulbright Student Program	2005	SRI International	Alumni from 14 countries who had participated between 1980 and 2000	14 countries	Quantitative: Online Survey (open-ended responses)
PROFESSIONAL EXCHANGE					
Peace Corps	2015	Office of Strategic Information, Research and Planning	Peace Corps Counterparts across 65 countries	Global; all Peace Corps countries (65)	Quantitative: Telephone survey
Peace Corps	2016	Office of Strategic Information, Research and Planning	Peace Corps Counterparts across 64 countries	Global; all Peace Corps countries (64)	Quantitative: Telephone survey
Peace Corps	2019	Office of Strategic Information, Research and Planning	All actively serving Peace Corps Volunteers (PCVs)	Global, all Peace Corps countries	Quantitative: Online survey
Peace Corps	2016	Office of Strategic Information, Research and Planning	25 analytical reports (2009–2014) based on the Host Country Impact Studies (HCIS's) conducted between 2008 and 2012	25 countries	Qualitative – meta-review of 25 existing studies

Fulbright English Teaching Assistant (ETA)	2011-2014	EurekaFacts, LLC	Fulbright ETAs who had participated between 2004-2005 or 2009-2010; Teachers and administrators at secondary schools, universities and other institutions where Fulbright ETAs were placed, as well as with Fulbright Commission staff, Ministry of Education officials, Regional English Language Officers, Embassy Public Affairs staff and other Embassy English language programming staff.	Chile, Thailand, Turkey and Russia.	Quantitative: Online Survey (open-ended responses)
African Women's Entrepreneurship Program (AWEP) - International Visitor Leadership Program (IVLP)	2015-2017	General Dynamics Information Technology (GDIT) & The District Communications Group; local research partner in Ghana	Participants from 2011-2015 across 48 Sub-Saharan countries; beneficiaries, experts and Embassy personnel	Sub-Saharan Africa; fieldwork in Ghana, Benin, Kenya, & Madagascar	Mixed methods: Telephone survey, In-depth interviews
U.S. Fulbright Scholar Program	1999-2002	SRI International	Stratified sample of 1,004 alumni who had participated between 1976 and 1999	Unknown	Quantitative
International Visitor Leadership Program (IVLP)	2006	ORC Macro; local research partner Institute for Comparative Social Research in Moscow	Participants from 1996 through 2001 from Georgia, Kazakhstan, Russia and Ukraine	Georgia, Kazakhstan, Russia and Ukraine	Qualitative: interviews and focus groups

Appendix C: Interview Protocol

1. To start off, can you please share with me a little of your background in evaluation?
    a. [Probe] How many years have you worked in evaluation?
    b. [Probe] What has been your role as an evaluator? (internal. Consultant, etc.)
    c. [Probe] Do you consider yourself to have a field of expertise in evaluation?
2. What process do you undertake when you conduct an evaluation?
    a. [Probe] How do you make decisions about what to evaluate?
    b. [Probe] What methodology do you prefer to use and why?
    c. [Probe] If applicable, how have you previously approached evaluating programs with goals that are challenging to measure? Soft Outcomes?
3. Have you ever worked on evaluating any programs funded by the United States Government?
    a. [Probe] Have you worked specifically on evaluations for any Educational or Cultural Exchange programs funded by the ECA or for Peace Corps?
    b. [Probe] What are some observations you made about working as an evaluator on those programs?
    c. [Probe] Were there any specifications or standards that you had to fulfill in your evaluation that were required by ECA or Peace Corps?
    d. [Probe] To the best of your knowledge, is there an existing framework for all evaluations that ECA or Peace Corps utilizes?
4. If you have worked on an evaluation for ECA or Peace Corps, what was the process you went through to conduct the evaluation?
    a. [Probe] How did you collect data?
    b. [Probe] How did you design the evaluation questions?
    c. [Probe] What program goals did you work to measure during this evaluation?
    d. [Probe] What were considerations you had in the design and facilitation of the evaluation?
    e. [Probe] Were there any challenges you encountered during this evaluation? If so, can you please describe?
5. Did this program have a goal of “mutual understanding”?
    a. [Probe] If so, can you please help define for me how the program facilitated gaining “mutual understanding” among program participants?
    b. [Probe] Did you try to measure this goal during your evaluation?
    c. [Probe] If so, what strategy did you utilize to measure “mutual understanding” for this program?
    d. [Probe] Do you have any ideas or suggestions for how you could have done it better or how you think it should be done in the future?
6. Do you have any final thoughts? Feel free to share anything further related to your expertise, background and experience as an evaluator that you think might be relevant.

Appendix D: Code Tree

Parent Code	Child Code I	Child Code II	Child Code III
Challenges	Budget		
	Bureaucracy	Management	
		Reporting Requirements	
	Congress		
	Competition		
	Donor requirements		
	Evaluation anxiety		
	Fieldwork contexts		
	Learning		
	Longitudinal		
	Methodology		
	Missing data		
	Mitigating bias		
	Policy Shifts		
	Response rates		
	Self-reported data		
	Staffing		
	Timing		
	Turnover		
Level of measurement	Long-term outcome		
	Output		
	Short-term outcome		
Methodology	Data collection strategies	Focus groups	
		Interviews	
		Online surveys	
		Disaggregation	
		Polls	
		Program records	
		Site visits	
	Evaluation Questions		
	Experimental (Control Group)		
	Mixed methods		
	Participatory		
	Sampling		
	Social Network Analysis (SNA)		
	Triangulation		

Parent Code	Child Code I	Child Code II	Child Code III
	Previous eval efforts		
Program Details	Alumni engagement		
	Goals	Mutual Understanding	
		Program-specific goal(s)	
	Participant demographics	Age	
		Country	
		Education level	
		Gender	
		Language	
		Region	
		Religion	
		Sector	
	Program Description		
	Program Name		
	Type	Academic Exchange	
		Cultural Exchange	
		Professional Exchange	
Recommendations			
Results & Impacts	Beneficiaries		
	Community-level	Capacity Building	
		Economic	
		Internationalization	
		Legal or Advocacy	
		Mutual understanding	
		Volunteering	
	Continued engagement		
	Individual-level	Confidence	
		Credibility	
		Educational gains	
		Employment gains	
		Increase in skills	Academic skills
			Job skills
			Language and communication skills
			Tech skills
		Knowledge	
		Leadership	

Parent Code	Child Code I	Child Code II	Child Code III
		Mutual understanding with USA	
		Network increases	
		Relationships with Americans	Advisors
			Cohort-Members
			Colleagues
			Community members
			Friends
			Host families
			Students
			US Embassy
		Relationships with other country nationals	Advisors
			Cohort-members
			Colleagues
			Community members
			Friends
			Host families
			Students
		New resources	
	Multiplier affect		
	Negative impacts		